arXiv:1001.2612v2 [math.OC] 11 May 2011

On distributed convex optimization under inequality and equality constraints via primal-dual subgradient methods

Minghui Zhu and Sonia Martínez

Abstract

We consider a general multi-agent convex optimization problem where the agents are to collectively minimize a global objective function subject to a global inequality constraint, a global equality constraint, and a global constraint set. The objective function is defined by a sum of local objective functions, while the global constraint set is produced by the intersection of local constraint sets. In particular, we study two cases: one where the equality constraint is absent, and the other where the local constraint sets are identical. We devise two distributed primal-dual subgradient algorithms which are based on the characterization of the primal-dual optimal solutions as the saddle points of the Lagrangian and penalty functions. These algorithms can be implemented over networks with changing topologies but satisfying a standard connectivity property, and allow the agents to asymptotically agree on optimal solutions and optimal values of the optimization problem under the Slater's condition.

I. Introduction

Recent advances in sensing, communication and computation technologies are challenging the way in which control mechanisms are designed for their efficient exploitation in a coordinated manner. This has motivated a wealth of algorithms for information processing, cooperative control, and optimization of large-scale networked multi-agent systems performing a variety of tasks. Due to a lack of a centralized authority, the proposed algorithms aim to be executed by individual agents through local actions, with the main feature of being robust to dynamic changes of network topologies.

The authors are with Department of Mechanical and Aerospace Engineering, University of California, San Diego, 9500 Gilman Dr, La Jolla CA, 92093, {mizhu,soniamd}@ucsd.edu

In this paper, we consider a general multi-agent optimization problem where the goal is to minimize a global objective function, given as a sum of local objective functions, subject to global constraints, which include an inequality constraint, an equality constraint and a (state) constraint set. Each local objective function is convex and only known to one particular agent. On the other hand, the inequality (resp. equality) constraint is given by a convex (resp. affine) function and known by all agents. Each node has its own convex constraint set, and the global constraint set is defined as their intersection. This problem is motivated by others in distributed estimation [24] [30], distributed source localization [28], network utility maximization [15], optimal flow control in power systems [26], [33] and optimal shape changes of mobile robots [9]. An important feature of the problem is that the objective and (or) constraint functions depend upon a global decision vector. This requires the design of distributed algorithms where, on the one hand, agents can align their decisions through a local information exchange and, on the other hand, the common decisions will coincide with an optimal solution and the optimal value.

Literature Review. In [2] and [32], the authors develop a general framework for parallel and distributed computation over a set of processors. Consensus problems, a class of canonical problems on networked multi-agent systems, have been intensively studied since then. A necessarily incomplete list of references includes [11], [25] tackling continuous-time consensus, [5], [12], [18] investigating discrete-time versions, and [17] where asynchronous implementation of consensus algorithms is discussed. The papers [6], [14], [31] treat randomized consensus via gossip communication, achieving consensus through quantized information and consensus over random graphs, respectively. The convergence rate of consensus algorithms is discussed, e.g., in [27], [34], and the author in [7] derives conditions to achieve different consensus values.

In robotics and control communities, convex optimization has been exploited to design algorithms coordinating mobile multi-agent systems. In [8], in order to increase the connectivity of a multi-agent system, a distributed supergradient-based algorithm is proposed to maximize the second smallest eigenvalue of the Laplacian matrix of the state dependent proximity graph of agents. In [9], optimal shape changes of mobile robots are achieved through second-order cone programming techniques. In [10], a target tracking problem is addressed by means of a generic semidefinite program where the constraints of network connectivity and full target coverage are articulated as linear-matrix inequalities. In [19], in order to attain the highest possible positioning accuracy for mobile robots, the authors express the covariance matrix of the pose errors as a

functional relation of measurement frequencies, and then formulate a optimal sensing problem as a convex programming of measurement frequencies.

The recent papers [21], [23] are the most relevant to our work. In [21], the authors solve a multi-agent unconstrained convex optimization problem through a novel combination of average consensus algorithms with subgradient methods. More recently, the paper [23] further takes local constraint sets into account. To deal with these constraints, the authors in [23] present an extension of their distributed subgradient algorithm, by projecting the original algorithm onto the local constraint sets. Two cases are solved in [23]: the first assumes that the network topologies can dynamically change and satisfy a periodic strong connectivity assumption (i.e., the union of the network topologies over a bounded period of time is strongly connected), but then the local constraint sets are identical; the second requires that the communication graphs are (fixed and) complete and then the local constraint sets can be different. Another related paper is [13] where a special case of [23], the network topology is fixed and all the local constraint sets are identical, is addressed.

Statement of Contributions. Building on the work [23], this paper further incorporates global inequality and equality constraints. More precisely, we study two cases: one in which the equality constraint is absent, and the other in which the local constraint sets are identical. For the first case, we adopt a Lagrangian relaxation approach, define a Lagrangian dual problem and devise the distributed Lagrangian primal-dual subgradient algorithm (DLPDS, for short) based on the characterization of the primal-dual optimal solutions as the saddle points of the Lagrangian function. The DLPDS algorithm involves each agent updating its estimates of the saddle points via a combination of an average consensus step, a subgradient (or supgradient) step and a primal (or dual) projection step onto its local constraint set (or a compact set containing the dual optimal set). The DLPDS algorithm is shown to asymptotically converge to a pair of primal-dual optimal solutions under the Slater's condition and the periodic strong connectivity assumption. Furthermore, each agent asymptotically agrees on the optimal value by implementing a dynamic average consensus algorithm developed in [35], which allows a multi-agent system to track time-varying average values.

For the second case, to dispense with the additional equality constraint, we adopt a penalty relaxation approach, while defining a penalty dual problem and devising the distributed penalty primal-dual subgradient algorithm (DPPDS, for short). Unlike the first case, the dual optimal

set of the second case may not be bounded, and thus the dual projection steps are not involved in the DPPDS algorithm. It renders that dual estimates and thus (primal) subgradients may not be uniformly bounded. This challenge is addressed by a more careful choice of step-sizes. We show that the DPPDS algorithm asymptotically converges to a primal optimal solution and the optimal value under the Slater's condition and the periodic strong connectivity assumption.

For the special case where the global inequality and equality constraints are not taken into account, this paper extends the results in [23] to a more general scenario where the network topologies satisfy the periodic strong connectivity assumption, and the local constraint sets can be different, while relaxing an interior-point condition requirement. We refer the readers to Section VI-D for additional information.

II. Problem formulation and assumptions

A. Problem formulation

Consider a network of agents labeled by $V := \{1, \dots, N\}$ that can only interact with each other through local communication. The objective of the multi-agent group is to cooperatively solve the following optimization problem:

$$\min_{x\in\mathbb{R}^n} f(x) := \sum_{i=1}^{N} f^{[i]}(x), \quad \text{s.t.} \quad g(x) \leq 0, \quad h(x) = 0, \quad x \in X := \cap_{i=1}^{N} X^{[i]}, \tag{1}$$

where $f^{[i]} : \mathbb{R}^n \to \mathbb{R}$ is the convex objective function of agent i, $X^{[i]} \subseteq \mathbb{R}^n$ is the compact and convex constraint set of agent i, and x is a global decision vector. Assume that $f^{[i]}$ and $X^{[i]}$ are only known by agent i, and probably different. The function $g : \mathbb{R}^n \to \mathbb{R}^m$ is known to all the agents with each component g_ℓ, for $\ell \in \{1, \dots, m\}$, being convex. The inequality $g(x) \leq 0$ is understood component-wise; i.e., $g_\ell(x) \leq 0$, for all $\ell \in \{1, \dots, m\}$, and represents a global inequality constraint. The function $h : \mathbb{R}^n \to \mathbb{R}^\nu$, defined as $h(x) := Ax - b$ with $A \in \mathbb{R}^{\nu\times n}$, represents a global equality constraint, and is known to all the agents. We denote $Y := \{x \in \mathbb{R}^n \mid g(x) \leq 0, \quad h(x) = 0\}$, and assume that the set of feasible points is non-empty; i.e., $X \cap Y \neq \emptyset$. Since X is compact and Y is closed, then we can deduce that $X \cap Y$ is compact. The convexity of $f^{[i]}$ implies that of f and thus f is continuous. In this way, the optimal value p^* of the problem (1) is finite and X^*, the set of primal optimal points, is non-empty. Throughout this paper, we suppose the following Slater's condition holds:

Assumption 2.1 (Slater's Condition): There exists a vector $\bar{x} \in X$ such that $g(\bar{x}) < 0$ and $h(\bar{x}) = 0$. And there exists a relative interior point $\tilde{x}$ of X, i.e., $\tilde{x} \in X$ and there exists an open sphere S centered at $\tilde{x}$ such that $S \cap \text{aff}(X) \subset X$ with $\text{aff}(X)$ being the affine hull of X, such that $h(\tilde{x}) = 0$.

Remark 2.1: In this paper, the quantities (e.g., functions, scalars and sets) associated with agent i will be indexed by the superscript $[i]$.

In this paper, we will study two particular cases of problem (1): one in which the global equality constraint $h(x) = 0$ is not included, and the other in which all the local constraint sets are identical. For the case where the constraint $h(x) = 0$ is absent, the Slater's condition 2.1 reduces to the existence of a vector $\bar{x} \in X$ such that $g(\bar{x}) < 0$.

B. Network model

We will consider that the multi-agent network operates synchronously. The topology of the network at time $k \geq 0$ will be represented by a directed weighted graph $\mathcal{G}(k) = (V, E(k), A(k))$ where $A(k) := [a_j^i(k)] \in \mathbb{R}^{N \times N}$ is the adjacency matrix with $a_j^i(k) \geq 0$ being the weight assigned to the edge (j, i) and $E(k) \subset V \times V \setminus \text{diag}(V)$ is the set of edges with non-zero weights $a_j^i(k)$. The in-neighbors of node i at time k are denoted by $\mathcal{N}^{[i]}(k) = \{j \in V \mid (j, i) \in E(k) \text{ and } j \neq i\}$. We here make the following assumptions on the network communication graphs, which are standard in the analysis of average consensus algorithms; e.g., see [25], [27], and distributed optimization in [21], [23].

Assumption 2.2 (Non-degeneracy): There exists a constant $\alpha > 0$ such that $a_i^i(k) \geq \alpha$, and $a_j^i(k)$, for $i \neq j$, satisfies $a_j^i(k) \in \{0\} \cup [\alpha,\ 1],\;$ for all $k \geq 0$.

Assumption 2.3 (Balanced Communication): [1] It holds that $\sum_{j=1}^N a_j^i(k) = 1$ for all $i \in V$ and $k \geq 0$, and $\sum_{i=1}^N a_j^i(k) = 1$ for all $j \in V$ and $k \geq 0$.

Assumption 2.4 (Periodical Strong Connectivity): There is a positive integer B such that, for all $k_0 \geq 0$, the directed graph $(V, \bigcup_{k=0}^{B-1} E(k_0 + k))$ is strongly connected.

C. Notion and notations

The following notion of saddle point plays a critical role in our paper.

[1] It is also referred to as double stochasticity.

Definition 2.1 (Saddle point): Consider a function $\phi : X \times M \to \mathbb{R}$ where X and M are non-empty subsets of $\mathbb{R}^{\bar{n}}$ and $\mathbb{R}^{\bar{m}}$. A pair of vectors $(x^*, \mu^*) \in X \times M$ is called a saddle point of ϕ over $X \times M$ if $\phi(x^*, \mu) \le \phi(x^*, \mu^*) \le \phi(x, \mu^*)$ hold for all $(x, \mu) \in X \times M$.

Remark 2.2: Equivalently, (x^*, μ^*) is a saddle point of ϕ over $X \times M$ if and only if $(x^*, \mu^*) \in X \times M$, and $\sup_{\mu \in M} \phi(x^*, \mu) \le \phi(x^*, \mu^*) \le \inf_{x \in X} \phi(x, \mu^*)$. •

In this paper, we do not assume the differentiability of $f^{[i]}$ and g_ℓ. At the points where the function is not differentiable, the subgradient plays the role of the gradient. For a given convex function $F : \mathbb{R}^{\bar{n}} \to \mathbb{R}$ and a point $\tilde{x} \in \mathbb{R}^{\bar{n}}$, a *subgradient* of the function F at $\tilde{x}$ is a vector $\mathcal{D}F(\tilde{x}) \in \mathbb{R}^{\bar{n}}$ such that the following subgradient inequality holds for any $x \in \mathbb{R}^{\bar{n}}$:

$$\mathcal{D}F(\tilde{x})^T (x - \tilde{x}) \le F(x) - F(\tilde{x}).$$

Similarly, for a given concave function $G : \mathbb{R}^{\bar{m}} \to \mathbb{R}$ and a point $\bar{\mu} \in \mathbb{R}^{\bar{m}}$, a *supgradient* of the function G at $\bar{\mu}$ is a vector $\mathcal{D}G(\bar{\mu}) \in \mathbb{R}^{\bar{m}}$ such that the following supgradient inequality holds for any $\mu \in \mathbb{R}^{\bar{m}}$:

$$\mathcal{D}G(\bar{\mu})^T (\mu - \bar{\mu}) \ge G(\mu) - G(\bar{\mu}).$$

Given a set S, we denote by $\mathrm{co}(S)$ its convex hull. We let the function $[\cdot]^+ : \mathbb{R}^{\bar{m}} \to \mathbb{R}^{\bar{m}}_{\ge 0}$ denote the projection operator onto the non-negative orthant in $\mathbb{R}^{\bar{m}}$. For any vector $c \in \mathbb{R}^{\bar{n}}$, we denote $|c| := (|c_1|, \cdots, |c_{\bar{n}}|)^T$, while $\| \cdot \|$ is the 2-norm in the Euclidean space.

III. Case (i): absence of equality constraint

In this section, we study the case of problem (1) where the equality constraint $h(x) = 0$ is absent; i.e., problem (1) becomes

$$\min_{x \in \mathbb{R}^n} \sum_{i=1}^{N} f^{[i]}(x), \quad \text{s.t.} \quad g(x) \le 0, \quad x \in \cap_{i=1}^{N} X^{[i]}. \tag{2}$$

We first provide some preliminaries, including a Lagrangian saddle-point characterization of problem (2) and finding a superset containing the Lagrangian dual optimal set of problem (2). After that, we present the distributed Lagrangian primal-dual subgradient algorithm and summarize its convergence properties.

A. Preliminaries

We here develop some preliminary results which are essential to the design of the distributed Lagrangian primal-dual subgradient algorithm.

1) A Lagrangian saddle-point characterization: Firstly, problem (2) is equivalent to

$$\min_{x\in\mathbb{R}^n} f(x), \quad \text{s.t.} \quad Ng(x) \leq 0, \quad x \in X,$$

with associated Lagrangian dual problem given by

$$\max_{\mu\in\mathbb{R}^m} q_L(\mu), \quad \text{s.t.} \quad \mu \geq 0.$$

Here, the Lagrangian dual function, $q_L : \mathbb{R}^m_{\geq 0} \to \mathbb{R}$, is defined as $q_L(\mu) := \inf_{x\in X} \mathcal{L}(x,\mu)$, where $\mathcal{L} : \mathbb{R}^n \times \mathbb{R}^m_{\geq 0} \to \mathbb{R}$ is the *Lagrangian function* $\mathcal{L}(x,\mu) = f(x) + N\mu^T g(x)$. We denote the Lagrangian dual optimal value of the Lagrangian dual problem by d_L^* and the set of Lagrangian dual optimal points by D_L^*. As is well-known, under the Slater's condition 2.1, the property of strong duality holds; i.e., $p^* = d_L^*$, and $D_L^* \neq \emptyset$. The following theorem is a standard result on Lagrangian duality stating that the primal and Lagrangian dual optimal solutions can be characterized as the saddle points of the Lagrangian function.

Theorem 3.1 (Lagrangian Saddle-point Theorem [3]): The pair of $(x^*, \mu^*) \in X \times \mathbb{R}^m_{\geq 0}$ is a saddle point of the Lagrangian function $\mathcal{L}$ over $X \times \mathbb{R}^m_{\geq 0}$ if and only if it is a pair of primal and Lagrangian dual optimal solutions and the following *Lagrangian minimax equality* holds:

$$\sup_{\mu\in\mathbb{R}^m_{\geq 0}} \inf_{x\in X} \mathcal{L}(x,\mu) = \inf_{x\in X} \sup_{\mu\in\mathbb{R}^m_{\geq 0}} \mathcal{L}(x,\mu).$$

This following lemma presents some preliminary analysis of Lagrangian saddle points.

Lemma 3.1 (Preliminary results of Lagrangian saddle points): Let M be any superset of D_L^*.

(a) If (x^*, μ^*) is a saddle point of $\mathcal{L}$ over $X \times \mathbb{R}^m_{\geq 0}$, then (x^*, μ^*) is also a saddle point of $\mathcal{L}$ over $X \times M$.

(b) There is at least one saddle point of $\mathcal{L}$ over $X \times M$.

(c) If $(\check{x}, \check{\mu})$ is a saddle point of $\mathcal{L}$ over $X \times M$, then $\mathcal{L}(\check{x}, \check{\mu}) = p^*$ and $\check{\mu}$ is Lagrangian dual optimal.

Proof: (a) It just follows from the definition of saddle point of $\mathcal{L}$ over $X \times M$.

(b) Observe that

$$\sup_{\mu\in\mathbb{R}^m_{\geq 0}} \inf_{x\in X} \mathcal{L}(x,\mu) = \sup_{\mu\in\mathbb{R}^m_{\geq 0}} q_L(\mu) = d_L^*,$$

$$\inf_{x\in X} \sup_{\mu\in\mathbb{R}^m_{\geq 0}} \mathcal{L}(x,\mu) = \inf_{x\in X\cap Y} f(x) = p^*.$$

Since the Slater's condition 2.1 implies zero duality gap, the Lagrangian minimax equality holds. From Theorem 3.1 it follows that the set of saddle points of $\mathcal{L}$ over $X \times \mathbb{R}^m_{\geq 0}$ is the Cartesian product $X^* \times D^*_L$. Recall that X^* and D^*_L are non-empty, so we can guarantee the existence of the saddle point of $\mathcal{L}$ over $X \times \mathbb{R}^m_{\geq 0}$. Then by (a), we have that (b) holds.

(c) Pick any saddle point (x^*, μ^*) of $\mathcal{L}$ over $X \times \mathbb{R}^m_{\geq 0}$. Since the Slater's condition 2.1 holds, from Theorem 3.1 one can deduce that (x^*, μ^*) is a pair of primal and Lagrangian dual optimal solutions. This implies that

$$d^*_L = \inf_{x \in X} \mathcal{L}(x, \mu^*) \leq \mathcal{L}(x^*, \mu^*) \leq \sup_{\mu \in \mathbb{R}^m_{\geq 0}} \mathcal{L}(x^*, \mu) = p^*.$$

From Theorem 3.1, we have $d^*_L = p^*$. Hence, $\mathcal{L}(x^*, \mu^*) = p^*$. On the other hand, we pick any saddle point $(\check{x}, \check{\mu})$ of $\mathcal{L}$ over $X \times M$. Then for all $x \in X$ and $\mu \in M$, it holds that $\mathcal{L}(\check{x}, \mu) \leq \mathcal{L}(\check{x}, \check{\mu}) \leq \mathcal{L}(x, \check{\mu})$. By Theorem 3.1, then $\mu^* \in D^*_L \subseteq M$. Recall $x^* \in X$, and thus we have $\mathcal{L}(\check{x}, \mu^*) \leq \mathcal{L}(\check{x}, \check{\mu}) \leq \mathcal{L}(x^*, \check{\mu})$. Since $\check{x} \in X$ and $\check{\mu} \in \mathbb{R}^m_{\geq 0}$, we have $\mathcal{L}(x^*, \check{\mu}) \leq \mathcal{L}(x^*, \mu^*) \leq \mathcal{L}(\check{x}, \mu^*)$. Combining the above two relations gives that $\mathcal{L}(\check{x}, \check{\mu}) = \mathcal{L}(x^*, \mu^*) = p^*$. From Remark 2.2 we see that $\mathcal{L}(\check{x}, \check{\mu}) \leq \inf_{x \in X} \mathcal{L}(x, \check{\mu}) = q_L(\check{\mu})$. Since $\mathcal{L}(\check{x}, \check{\mu}) = p^* = d^*_L \geq q_L(\check{\mu})$, then $q_L(\check{\mu}) = d^*_L$ and thus $\check{\mu}$ is a Lagrangian dual optimal solution. ■

Remark 3.1: Despite that (c) holds, the reverse of (a) may not be true in general. In particular, x^* may be infeasible; i.e., $g_\ell(x^*) > 0$ for some $\ell \in \{1, \ldots, m\}$. •

2) A upper estimate of the Lagrangian dual optimal set: In what follows, we will find a compact superset of D^*_L. To do that, we define the following primal problem for each agent i:

$$\min_{x \in \mathbb{R}^n} f^{[i]}(x), \quad \text{s.t.} \quad g(x) \leq 0, \quad x \in X^{[i]}.$$

Due to the fact that $X^{[i]}$ is compact and the $f^{[i]}$ are continuous, the primal optimal value p^*_i of each agent's primal problem is finite and the set of its primal optimal solutions is non-empty. The associated dual problem is given by

$$\max_{\mu \in \mathbb{R}^m} q^{[i]}(\mu), \quad \text{s.t.} \quad \mu \geq 0.$$

Here, the dual function $q^{[i]} : \mathbb{R}^m_{\geq 0} \to \mathbb{R}$ is defined by $q^{[i]}(\mu) := \inf_{x \in X^{[i]}} \mathcal{L}^{[i]}(x, \mu)$, where $\mathcal{L}^{[i]} : \mathbb{R}^n \times \mathbb{R}^m_{\geq 0} \to \mathbb{R}$ is the Lagrangian function of agent i and given by $\mathcal{L}^{[i]}(x, \mu) = f^{[i]}(x) + \mu^T g(x)$. The corresponding dual optimal value is denoted by d^*_i. In this way, $\mathcal{L}$ is decomposed into a sum of local Lagrangian functions; i.e., $\mathcal{L}(x, \mu) = \sum_{i=1}^N \mathcal{L}^{[i]}(x, \mu)$.

Define now the set-valued map $Q : \mathbb{R}^m_{\geq 0} \to 2^{(\mathbb{R}^m_{\geq 0})}$ by $Q(\tilde{\mu}) = \{\mu \in \mathbb{R}^m_{\geq 0} \mid q_L(\mu) \geq q_L(\tilde{\mu})\}$. Additionally, define a function $\gamma : X \to \mathbb{R}$ by $\gamma(x) = \min_{\ell \in \{1,\ldots,m\}}\{-g_\ell(x)\}$. Observe that if x is a Slater vector, then $\gamma(x) > 0$. The following lemma is a direct result of Lemma 1 in [20].

Lemma 3.2 (Boundedness of dual solution sets): The set $Q(\tilde{\mu})$ is bounded for any $\tilde{\mu} \in \mathbb{R}^m_{\geq 0}$, and, in particular, for any Slater vector $\bar{x}$, it holds that $\max_{\mu \in Q(\tilde{\mu})} \|\mu\| \leq \frac{1}{\gamma(\bar{x})}(f(\bar{x}) - q_L(\tilde{\mu}))$. □

Notice that $D_L^* = \{\mu \in \mathbb{R}^m_{\geq 0} \mid q_L(\mu) \geq d_L^*\}$. Picking any Slater vector $\bar{x} \in X$, and letting $\tilde{\mu} = \mu^* \in D_L^*$ in Lemma 3.2 gives that

$$\max_{\mu^* \in D_L^*} \|\mu^*\| \leq \frac{1}{\gamma(\bar{x})}(f(\bar{x}) - d_L^*). \tag{3}$$

Define the function $r : X \times \mathbb{R}^m_{\geq 0} \to \mathbb{R} \cup \{+\infty\}$ by $r(x, \mu) := \frac{N}{\gamma(x)} \max_{i \in V}\{f^{[i]}(x) - q^{[i]}(\mu)\}$. By the property of weak duality, it holds that $d_i^* \leq p_i^*$ and thus $f^{[i]}(x) \geq q^{[i]}(\mu)$ for any $(x, \mu) \in X \times \mathbb{R}^m_{\geq 0}$. Since $\gamma(\bar{x}) > 0$, thus $r(\bar{x}, \mu) \geq 0$ for any $\mu \in \mathbb{R}^m_{\geq 0}$. With this observation, we pick any $\tilde{\mu} \in \mathbb{R}^m_{\geq 0}$ and the following set is well-defined: $\bar{M}^{[i]}(\bar{x}, \tilde{\mu}) := \{\mu \in \mathbb{R}^m_{\geq 0} \mid \|\mu\| \leq r(\bar{x}, \tilde{\mu}) + \theta^{[i]}\}$ for some $\theta^{[i]} \in \mathbb{R}_{>0}$. Observe that for all $\mu \in \mathbb{R}^m_{\geq 0}$:

$$q_L(\mu) = \inf_{x \in \cap_{i=1}^N X^{[i]}} \sum_{i=1}^N (f^{[i]}(x) + \mu^T g(x)) \geq \sum_{i=1}^N \inf_{x \in X^{[i]}} (f^{[i]}(x) + \mu^T g(x)) = \sum_{i=1}^N q^{[i]}(\mu). \tag{4}$$

Since $d_L^* \geq q_L(\tilde{\mu})$, it follows from (3) and (4) that

$$\begin{aligned}\max_{\mu^* \in D_L^*} \|\mu^*\| &\leq \frac{1}{\gamma(\bar{x})}(f(\bar{x}) - q_L(\tilde{\mu})) \leq \frac{1}{\gamma(\bar{x})}(f(\bar{x}) - \sum_{i=1}^N q^{[i]}(\tilde{\mu})) \\ &\leq \frac{N}{\gamma(\bar{x})} \max_{i \in V}\{f^{[i]}(\bar{x}) - q^{[i]}(\tilde{\mu})\} = r(\bar{x}, \tilde{\mu}).\end{aligned}$$

Hence, we have $D_L^* \subseteq \bar{M}^{[i]}(\bar{x}, \tilde{\mu})$ for all $i \in V$.

Note that in order to compute $\bar{M}^{[i]}(\bar{x}, \tilde{\mu})$, all the agents have to agree on a common Slater vector $\bar{x} \in \cap_{i=1}^N X^{[i]}$ which should be obtained in a distributed fashion. To handle this difficulty, we now propose a distributed algorithm, namely *Distributed Slater-vector Computation Algorithm*, which allows each agent i to compute a superset of $\bar{M}^{[i]}(\bar{x}, \tilde{\mu})$.

Initially, each agent i chooses a common value $\tilde{\mu} \in \mathbb{R}^m_{\geq 0}$; e.g., $\tilde{\mu} = 0$, and computes two positive constants $b^{[i]}(0)$ and $c^{[i]}(0)$ such that $b^{[i]}(0) \geq \sup_{x \in J^{[i]}}\{f^{[i]}(x) - q^{[i]}(\tilde{\mu})\}$ and $c^{[i]}(0) \leq \min_{1 \leq \ell \leq m} \inf_{x \in J^{[i]}}\{-g_\ell(x)\}$ where $J^{[i]} := \{x \in X^{[i]} \mid g(x) < 0\}$.

At every time $k \geq 0$, each agent i updates its estimates by using the following rules:

$$b^{[i]}(k+1) = \max_{j\in\mathcal{N}^{[i]}(k)\cup\{i\}} b^{[j]}(k), \quad c^{[i]}(k+1) = \min_{j\in\mathcal{N}^{[i]}(k)\cup\{i\}} c^{[j]}(k).$$

Lemma 3.3 (Convergence properties of the distributed Slater-vector Computation Algorithm): Assume that the periodical strong connectivity assumption 2.4 holds. Consider the sequences of $\{b^{[i]}(k)\}$ and $\{c^{[i]}(k)\}$ generated by the Distributed Slater-vector Computation Algorithm. It holds that after at most $(N-1)B$ steps, all the agents reach the consensus, i.e., $b^{[i]}(k) = b^* := \max_{j\in V} b^{[j]}(0)$ and $c^{[i]}(k) = c^* := \min_{j\in V} c^{[j]}(0)$ for all $k \geq (N-1)B$. Furthermore, we have $M^{[i]}(\tilde{\mu}) := \{\mu \in \mathbb{R}^m_{\geq 0} \mid \|\mu\| \leq \frac{Nb^*}{c^*} + \theta^{[i]}\} \supseteq \bar{M}^{[i]}(\bar{x},\tilde{\mu})$ for $i \in V$.

Proof: It is not difficult to verify achieving max-consensus and min-consensus by using the periodical strong connectivity assumption 2.4. Note that $J := \{x \in X \mid g(x) < 0\} \subseteq J^{[i]}$, $\forall i \in V$. Hence, we have

$$\max_{i\in V}\sup_{x\in J}\{f^{[i]}(x) - q^{[i]}(\tilde{\mu})\} \leq \max_{i\in V}\sup_{x\in J^{[i]}}\{f^{[i]}(x) - q^{[i]}(\tilde{\mu})\} \leq b^*,$$

$$\inf_{x\in J}\min_{1\leq\ell\leq m}\{-g_\ell(x)\} \geq \min_{i\in V}\inf_{x\in J^{[i]}}\min_{1\leq\ell\leq m}\{-g_\ell(x)\} \geq c^*.$$

Since $\bar{x} \in J$, then the following estimate on $r(\bar{x},\tilde{\mu})$ holds:

$$r(\bar{x},\tilde{\mu}) \leq \frac{N\sup_{x\in J}\max_{i\in V}\{f^{[i]}(x) - q^{[i]}(\tilde{\mu})\}}{\inf_{x\in J}\min_{1\leq\ell\leq m}\{-g_\ell(x)\}} \leq \frac{Nb^*}{c^*}.$$

The desired result immediately follows. ■

From Lemma 3.3 and the fact that $D^*_L \subseteq \bar{M}^{[i]}(\bar{x},\tilde{\mu})$, we can see that the set of $M(\tilde{\mu}) := \cap_{i=1}^N M^{[i]}(\tilde{\mu})$ contains D^*_L. In addition, $M^{[i]}(\tilde{\mu})$ and $M(\tilde{\mu})$ are non-empty, compact and convex. To simplify the notations, we will use the shorthands $M^{[i]} := M^{[i]}(\tilde{\mu})$ and $M := M(\tilde{\mu})$.

3) Convexity of $\mathcal{L}$: For each $\mu \in \mathbb{R}^m_{\geq 0}$, we define the function $\mathcal{L}^{[i]}_\mu : \mathbb{R}^n \to \mathbb{R}$ as $\mathcal{L}^{[i]}_\mu(x) := \mathcal{L}^{[i]}(x,\mu)$. Note that $\mathcal{L}^{[i]}_\mu$ is convex since it is a nonnegative weighted sum of convex functions. For each $x \in \mathbb{R}^n$, we define the function $\mathcal{L}^{[i]}_x : \mathbb{R}^m_{\geq 0} \to \mathbb{R}$ as $\mathcal{L}^{[i]}_x(\mu) := \mathcal{L}^{[i]}(x,\mu)$. It is easy to check that $\mathcal{L}^{[i]}_x$ is a concave (actually affine) function. Then the Lagrangian function $\mathcal{L}$ is the sum of a collection of convex-concave local functions. This property motivates us to significantly extend primal-dual subgradient methods in [1], [22] to the networked multi-agent scenario.

B. Distributed Lagrangian primal-dual subgradient algorithm

Here, we introduce the *Distributed Lagrangian Primal-Dual Subgradient* Algorithm (DLPDS, for short) to find a saddle point of the Lagrangian function $\mathcal{L}$ over $X \times M$ and the optimal value.

This saddle point will coincide with a pair of primal and Lagrangian dual optimal solutions which is not always the case; see Remark 3.1.

Through the algorithm, at each time k, each agent i maintains the estimate of $(x^{[i]}(k), \mu^{[i]}(k))$ to the saddle point of the Lagrangian function $\mathcal{L}$ over $X \times M$ and the estimate of $y^{[i]}(k)$ to p^*. To produce $x^{[i]}(k+1)$ (resp. $\mu^{[i]}(k+1)$), agent i takes a convex combination $v_x^{[i]}(k)$ (resp. $v_\mu^{[i]}(k)$) of its estimate $x^{[i]}(k)$ (resp. $\mu^{[i]}(k)$) with the estimates sent from its neighboring agents at time k, makes a subgradient (resp. supgradient) step to minimize (resp. maximize) the local Lagrangian function $\mathcal{L}^{[i]}$, and takes a primal (resp. dual) projection onto the local constraint $X^{[i]}$ (resp. $M^{[i]}$). Furthermore, agent i generates the estimate $y^{[i]}(k+1)$ by taking a convex combination $v_y^{[i]}(k)$ of its estimate $y^{[i]}(k)$ with the estimates of its neighbors at time k and taking one step to track the variation of the local objective function $f^{[i]}$. The DLPDS algorithm is formally stated as follows:

Initially, each agent i picks a common $\tilde{\mu} \in \mathbb{R}^m_{\geq 0}$ and computes the set $M^{[i]}$ with some $\theta^{[i]} > 0$ by using the Distributed Slater-vector Computation Algorithm. Furthermore, agent i chooses any initial state $x^{[i]}(0) \in X^{[i]}$, $\mu^{[i]}(0) \in \mathbb{R}^m_{\geq 0}$, and $y^{[i]}(1) = N f^{[i]}(x^{[i]}(0))$.

At every $k \geq 0$, each agent i generates $x^{[i]}(k+1)$, $\mu^{[i]}(k+1)$ and $y^{[i]}(k+1)$ according to the following rules:

$$v_x^{[i]}(k) = \sum_{j=1}^N a_j^i(k) x^{[j]}(k), \quad v_\mu^{[i]}(k) = \sum_{j=1}^N a_j^i(k) \mu^{[j]}(k), \quad v_y^{[i]}(k) = \sum_{j=1}^N a_j^i(k) y^{[j]}(k),$$

$$x^{[i]}(k+1) = P_{X^{[i]}}[v_x^{[i]}(k) - \alpha(k)\mathcal{D}_x^{[i]}(k)], \quad \mu^{[i]}(k+1) = P_{M^{[i]}}[v_\mu^{[i]}(k) + \alpha(k)\mathcal{D}_\mu^{[i]}(k)],$$

$$y^{[i]}(k+1) = v_y^{[i]}(k) + N(f^{[i]}(x^{[i]}(k)) - f^{[i]}(x^{[i]}(k-1))),$$

where $P_{X^{[i]}}$ (resp. $P_{M^{[i]}}$) is the projection operator onto the set $X^{[i]}$ (resp. $M^{[i]}$), the scalars $a_j^i(k)$ are non-negative weights and the scalars $\alpha(k) > 0$ are step-sizes[2]. We use the shorthands $\mathcal{D}_x^{[i]}(k) \equiv \mathcal{D}\mathcal{L}^{[i]}_{v_\mu^{[i]}(k)}(v_x^{[i]}(k))$, and $\mathcal{D}_\mu^{[i]}(k) \equiv \mathcal{D}\mathcal{L}^{[i]}_{v_x^{[i]}(k)}(v_\mu^{[i]}(k))$.

The following theorem summarizes the convergence properties of the DLPDS algorithm where agents asymptotically agree upon a pair of primal-dual optimal solutions.

Theorem 3.2 (Convergence properties of the DLPDS algorithm): Consider the optimization problem (2). Let the non-degeneracy assumption 2.2, the balanced communication assumption 2.3 and the periodic strong connectivity assumptions 2.4 hold. Consider the sequences of

[2]Each agent i executes the update law of $y^{[i]}(k)$ for $k \geq 1$.

$\{x^{[i]}(k)\}$, $\{\mu^{[i]}(k)\}$ and $\{y^{[i]}(k)\}$ of the distributed Lagrangian primal-dual subgradient algorithm with the step-sizes $\{\alpha(k)\}$ satisfying $\lim_{k\to+\infty}\alpha(k)=0$, $\sum_{k=0}^{+\infty}\alpha(k)=+\infty$, and $\sum_{k=0}^{+\infty}\alpha(k)^2<+\infty$. Then, there is a pair of primal and Lagrangian dual optimal solutions $(x^*,\mu^*)\in X^*\times D_L^*$ such that $\lim_{k\to+\infty}\|x^{[i]}(k)-x^*\|=0$ and $\lim_{k\to+\infty}\|\mu^{[i]}(k)-\mu^*\|=0$ for all $i\in V$. Furthermore, we have that $\lim_{k\to+\infty}\|y^{[i]}(k)-p^*\|=0$ for all $i\in V$.

Remark 3.2: For a convex-concave function, continuous-time gradient-based methods are proved in [1] to converge globally towards the saddle-point. Recently, [22] presents (discrete-time) primal-dual subgradient methods which relax the differentiability in [1] and further incorporate state constraints. The method in [1] is adopted by [16] and [29] to study a distributed optimization problem on fixed graphs where objective functions are separable.

The DLPDS algorithm is a generalization of primal-dual subgradient methods in [22] to the networked multi-agent scenario. It is also an extension of the distributed projected subgradient algorithm in [23] to solve multi-agent convex optimization problems with inequality constraints. Additionally, the DLPDS algorithm enables agents to find the optimal value. Furthermore, the DLPDS algorithm objective is that of reaching a saddle point of the Lagrangian function in contrast to achieving a (primal) optimal solution in [23]. •

IV. CASE (II): IDENTICAL LOCAL CONSTRAINT SETS

In last section, we study the case where the equality constraint is absent in problem (1). In this section, we turn our attention to another case of problem (1) where $h(x)=0$ is taken into account but we require that local constraint sets are identical; i.e., $X^{[i]}=X$ for all $i\in V$. We first adopt a penalty relaxation and provide a penalty saddle-point characterization of primal problem (1) with $X^{[i]}=X$. We then introduce the distributed penalty primal-dual subgradient algorithm, followed by its convergence properties and some remarks.

A. Preliminaries

Some preliminary results are presented in this part, and these results are essential to the development of the distributed penalty primal-dual subgradient algorithm.

1) A penalty saddle-point characterization: Note that the primal problem (1) with $X^{[i]} = X$ is trivially equivalent to the following:

$$\min_{x\in\mathbb{R}^n} f(x), \quad \text{s.t.} \quad Ng(x) \leq 0, \quad Nh(x) = 0, \quad x \in X, \tag{5}$$

with associated penalty dual problem given by

$$\max_{\mu\in\mathbb{R}^m,\lambda\in\mathbb{R}^\nu} q_P(\mu,\lambda), \quad \text{s.t.} \quad \mu \geq 0, \quad \lambda \geq 0. \tag{6}$$

Here, the penalty dual function, $q_P : \mathbb{R}^m_{\geq 0}\times\mathbb{R}^\nu_{\geq 0} \to \mathbb{R}$, is defined by $q_P(\mu,\lambda) := \inf_{x\in X} \mathcal{H}(x,\mu,\lambda)$, where $\mathcal{H} : \mathbb{R}^n \times \mathbb{R}^m_{\geq 0} \times \mathbb{R}^\nu_{\geq 0} \to \mathbb{R}$ is the *penalty function* given by $\mathcal{H}(x,\mu,\lambda) = f(x) + N\mu^T[g(x)]^+ + N\lambda^T|h(x)|$. We denote the penalty dual optimal value by d^*_P and the set of penalty dual optimal solutions by D^*_P. We define the penalty function $\mathcal{H}^{[i]}(x,\mu,\lambda) : \mathbb{R}^n\times\mathbb{R}^m_{\geq 0}\times\mathbb{R}^\nu_{\geq 0} \to \mathbb{R}$ for each agent i as follows: $\mathcal{H}^{[i]}(x,\mu,\lambda) = f^{[i]}(x) + \mu^T[g(x)]^+ + \lambda^T|h(x)|$. In this way, we have that $\mathcal{H}(x,\mu,\lambda) = \sum_{i=1}^N \mathcal{H}^{[i]}(x,\mu,\lambda)$. As proven in the next lemma, the Slater's condition 2.1 ensures zero duality gap and the existence of penalty dual optimal solutions.

Lemma 4.1 (Strong duality and non-emptyness of the penalty dual optimal set): The values of p^* and d^*_P coincide, and D^*_P is non-empty.

Proof: Consider the auxiliary Lagrangian function $\mathcal{L}_a : \mathbb{R}^n \times \mathbb{R}^m_{\geq 0} \times \mathbb{R}^\nu \to \mathbb{R}$ given by $\mathcal{L}_a(x,\mu,\lambda) = f(x) + N\mu^T g(x) + N\lambda^T h(x)$, with the associated dual problem defined by

$$\max_{\mu\in\mathbb{R}^m,\lambda\in\mathbb{R}^\nu} q_a(\mu,\lambda), \quad \text{s.t.} \quad \mu \geq 0. \tag{7}$$

Here, the dual function, $q_a : \mathbb{R}^m_{\geq 0} \times \mathbb{R}^\nu \to \mathbb{R}$, is defined by $q_a(\mu,\lambda) := \inf_{x\in X} \mathcal{L}_a(x,\mu,\lambda)$. The dual optimal value of problem (7) is denoted by d^*_a and the set of dual optimal solutions is denoted D^*_a. Since X is convex, f and g_ℓ, for $\ell \in \{1,\dots,m\}$, are convex, p^* is finite and the Slater's condition 2.1 holds, it follows from Proposition 5.3.5 in [3] that $p^* = d^*_a$ and $D^*_a \neq \emptyset$. We now proceed to characterize d^*_P and D^*_P. Pick any $(\mu^*,\lambda^*) \in D^*_a$. Since $\mu^* \geq 0$, then

$$\begin{aligned} d^*_a = q_a(\mu^*,\lambda^*) &= \inf_{x\in X}\{f(x) + N(\mu^*)^T g(x) + N(\lambda^*)^T h(x)\} \\ &\leq \inf_{x\in X}\{f(x) + N(\mu^*)^T[g(x)]^+ + N|\lambda^*|^T|h(x)|\} = q_P(\mu^*,|\lambda^*|) \leq d^*_P. \end{aligned} \tag{8}$$

On the other hand, pick any $x^* \in X^*$. Then x^* is feasible, i.e., $x^* \in X$, $[g(x^*)]^+ = 0$ and $|h(x^*)| = 0$. It implies that $q_P(\mu,\lambda) \leq \mathcal{H}(x^*,\mu,\lambda) = f(x^*) = p^*$ holds for any $\mu \in \mathbb{R}^m_{\geq 0}$ and $\lambda \in \mathbb{R}^\nu_{\geq 0}$, and thus $d^*_P = \sup_{\mu\in\mathbb{R}^m_{\geq 0},\lambda\in\mathbb{R}^\nu_{\geq 0}} q_P(\mu,\lambda) \leq p^* = d^*_a$. Therefore, we have $d^*_P = p^*$.

To prove the emptyness of D_P^*, we pick any $(\mu^*, \lambda^*) \in D_a^*$. From (8) and $d_a^* = d_P^*$, we can see that $(\mu^*, |\lambda^*|) \in D_P^*$ and thus $D_P^* \neq \emptyset$. ■

The following is a slight extension of Theorem 3.1 to penalty functions.

Theorem 4.1 (Penalty Saddle-point Theorem): The pair of (x^*, μ^*, λ^*) is a saddle point of the penalty function $\mathcal{H}$ over $X \times \mathbb{R}^m_{\geq 0} \times \mathbb{R}^\nu_{\geq 0}$ if and only if it is a pair of primal and penalty dual optimal solutions and the following *penalty minimax equality* holds:

$$\sup_{(\mu,\lambda)\in\mathbb{R}^m_{\geq 0}\times\mathbb{R}^\nu_{\geq 0}} \inf_{x\in X} \mathcal{H}(x,\mu,\lambda) = \inf_{x\in X} \sup_{(\mu,\lambda)\in\mathbb{R}^m_{\geq 0}\times\mathbb{R}^\nu_{\geq 0}} \mathcal{H}(x,\mu,\lambda).$$

Proof: The proof is analogous to that of Proposition 6.2.4 in [4], and for the sake of completeness, we provide the details here. It follows from Proposition 2.6.1 in [4] that (x^*, μ^*, λ^*) is a saddle point of $\mathcal{H}$ over $X \times \mathbb{R}^m_{\geq 0} \times \mathbb{R}^\nu_{\geq 0}$ if and only if the penalty minimax equality holds and the following conditions are satisfied:

$$\sup_{(\mu,\lambda)\in\mathbb{R}^m_{\geq 0}\times\mathbb{R}^\nu_{\geq 0}} \mathcal{H}(x^*,\mu,\lambda) = \min_{x\in X}\{\sup_{(\mu,\lambda)\in\mathbb{R}^m_{\geq 0}\times\mathbb{R}^\nu_{\geq 0}} \mathcal{H}(x,\mu,\lambda)\}, \tag{9}$$

$$\inf_{x\in X} \mathcal{H}(x,\mu^*,\lambda^*) = \max_{(\mu,\lambda)\in\mathbb{R}^m_{\geq 0}\times\mathbb{R}^\nu_{\geq 0}} \{\inf_{x\in X} \mathcal{H}(x,\mu,\lambda)\}. \tag{10}$$

Notice that $\inf_{x\in X} \mathcal{H}(x,\mu,\lambda) = q_P(\mu,\lambda)$; and if $x \in Y$, then $\sup_{(\mu,\lambda)\in\mathbb{R}^m_{\geq 0}\times\mathbb{R}^\nu_{\geq 0}} \mathcal{H}(x,\mu,\lambda) = f(x)$, otherwise, $\sup_{(\mu,\lambda)\in\mathbb{R}^m_{\geq 0}\times\mathbb{R}^\nu_{\geq 0}} \mathcal{H}(x,\mu,\lambda) = +\infty$. Hence, the penalty minimax equality is equivalent to $d_P^* = p^*$. Condition (9) is equivalent to the fact that x^* is primal optimal, and condition (10) is equivalent to (μ^*, λ^*) being a penalty dual optimal solution. ■

2) Convexity of $\mathcal{H}$: Since g_ℓ is convex and $[\cdot]^+$ is convex and non-decreasing, thus $[g_\ell(x)]^+$ is convex in x for each $\ell \in \{1,\ldots,m\}$. Denote $A := (a_1^T, \cdots, a_\nu^T)^T$. Since $|\cdot|$ is convex and $a_\ell^T x - b_\ell$ is an affine mapping, then $|a_\ell^T x - b_\ell|$ is convex in x for each $\ell \in \{1,\ldots,\nu\}$.

We denote $w := (\mu,\lambda)$. For each $w \in \mathbb{R}^m_{\geq 0} \times \mathbb{R}^\nu_{\geq 0}$, we define the function $\mathcal{H}_w^{[i]} : \mathbb{R}^n \to \mathbb{R}$ as $\mathcal{H}_w^{[i]}(x) := \mathcal{H}^{[i]}(x,w)$. Note that $\mathcal{H}_w^{[i]}(x)$ is convex in x by using the fact that a nonnegative weighted sum of convex functions is convex. For each $x \in \mathbb{R}^n$, we define the function $\mathcal{H}_x^{[i]} : \mathbb{R}^m_{\geq 0} \times \mathbb{R}^\nu_{\geq 0} \to \mathbb{R}$ as $\mathcal{H}_x^{[i]}(w) := \mathcal{H}^{[i]}(x,w)$. It is easy to check that $\mathcal{H}_x^{[i]}(w)$ is concave (actually affine) in w. Then the penalty function $\mathcal{H}(x,w)$ is the sum of convex-concave local functions.

Remark 4.1: The Lagrangian relaxation does not fit to our approach here since the Lagrangian function is not convex in x by allowing λ entries to be negative. •

B. Distributed penalty primal-dual subgradient algorithm

We are now in the position to devise the *Distributed Penalty Primal-Dual Subgradient* Algorithm (DPPDS, for short), that is based on the penalty saddle-point theorem 4.1, to find the optimal value and a primal optimal solution to primal problem (1) with $X^{[i]} = X$. The DPPDS algorithm is formally described as follow.

Initially, agent i chooses any initial state $x^{[i]}(0) \in X$, $\mu^{[i]}(0) \in \mathbb{R}^m_{\geq 0}$, $\lambda^{[i]}(0) \in \mathbb{R}^\nu_{\geq 0}$, and $y^{[i]}(1) = Nf^{[i]}(x^{[i]}(0))$. At every time $k \geq 0$, each agent i computes the following convex combinations:

$$v_x^{[i]}(k) = \sum_{j=1}^N a_j^i(k) x^{[j]}(k), \quad v_y^{[i]}(k) = \sum_{j=1}^N a_j^i(k) y^{[j]}(k),$$
$$v_\mu^{[i]}(k) = \sum_{j=1}^N a_j^i(k) \mu^{[j]}(k), \quad v_\lambda^{[i]}(k) = \sum_{j=1}^N a_j^i(k) \lambda^{[j]}(k),$$

and generates $x^{[i]}(k+1)$, $y^{[i]}(k+1)$, $\mu^{[i]}(k+1)$ and $\lambda^{[i]}(k+1)$ according to the following rules:

$$x^{[i]}(k+1) = P_X[v_x^{[i]}(k) - \alpha(k)\mathcal{S}_x^{[i]}(k)], \quad y^{[i]}(k+1) = v_y^{[i]}(k) + N(f^{[i]}(x^{[i]}(k)) - f^{[i]}(x^{[i]}(k-1))),$$
$$\mu^{[i]}(k+1) = v_\mu^{[i]}(k) + \alpha(k)[g(v_x^{[i]}(k))]^+, \quad \lambda^{[i]}(k+1) = v_\lambda^{[i]}(k) + \alpha(k)|h(v_x^{[i]}(k))|,$$

where P_X is the projection operator onto the set X, the scalars $a_j^i(k)$ are non-negative weights and the positive scalars $\{\alpha(k)\}$ are step-sizes[3]. The vector

$$\mathcal{S}_x^{[i]}(k) := \mathcal{D}f^{[i]}(v_x^{[i]}(k)) + \sum_{\ell=1}^m v_\mu^{[i]}(k)_\ell \mathcal{D}[g_\ell(v_x^{[i]}(k))]^+ + \sum_{\ell=1}^\nu v_\lambda^{[i]}(k)_\ell \mathcal{D}|h_\ell|(v_x^{[i]}(k))$$

is a subgradient of $\mathcal{H}^{[i]}_{w^{[i]}(k)}(x)$ at $x = v_x^{[i]}(k)$ where $w^{[i]}(k) := (v_\mu^{[i]}(k), v_\lambda^{[i]}(k))$ is the convex combination of dual estimates of agent i and its neighbors'.

Given a step-size sequence $\{\alpha(k)\}$, we define its sum over $[0,k]$ by $s(k) := \sum_{\ell=0}^k \alpha(\ell)$ and assume that:

Assumption 4.1 (Step-size assumption): The step-sizes satisfy $\lim\limits_{k\to+\infty} \alpha(k) = 0$, $\sum_{k=0}^{+\infty} \alpha(k) = +\infty$, $\sum_{k=0}^{+\infty} \alpha(k)^2 < +\infty$, and $\lim\limits_{k\to+\infty} \alpha(k+1)s(k) = 0$, $\sum_{k=0}^{+\infty} \alpha(k+1)^2 s(k) < +\infty$, $\sum_{k=0}^{+\infty} \alpha(k+1)^2 s(k)^2 < +\infty$.

[3]Each agent i executes the update law of $y^{[i]}(k)$ for $k \geq 1$.

The following theorem is the main result of this section, characterizing convergence properties of the DPPDS algorithm where a optimal solution and the optimal value are asymptotically agreed upon.

Theorem 4.2 (Convergence properties of the DPPDS algorithm): Consider the problem (1) with $X^{[i]} = X$. Let the non-degeneracy assumption 2.2, the balanced communication assumption 2.3 and the periodic strong connectivity assumption 2.4 hold. Consider the sequences of $\{x^{[i]}(k)\}$ and $\{y^{[i]}(k)\}$ of the distributed penalty primal-dual subgradient algorithm where the step-sizes $\{\alpha(k)\}$ satisfy the step-size assumption 4.1. Then there exists a primal optimal solution $\tilde{x} \in X^*$ such that $\lim_{k\to+\infty} \|x^{[i]}(k) - \tilde{x}\| = 0$ for all $i \in V$. Furthermore, we have $\lim_{k\to+\infty} \|y^{[i]}(k) - p^*\| = 0$ for all $i \in V$.

We here provide some remarks to conclude this section,.

Remark 4.2: As primal-dual (sub)gradient algorithm in [1], [22], the DPPDS algorithm produces a pair of primal and dual estimates at each step. Main differences include: firstly, the DPPDS algorithm extends the primal-dual subgradient algorithm in [22] to the multi-agent scenario; secondly, it further takes the equality constraint into account. The presence of the equality constraint can make D_P^* unbounded. Therefore, unlike the DLPDS algorithm, the DPPDS algorithm does not involve the dual projection steps onto compact sets. This may cause the subgradient $\mathcal{S}_x^{[i]}(k)$ not to be uniformly bounded, while the boundedness of subgradients is a standard assumption in the analysis of subgradient methods, e.g., see [3], [4], [20], [21], [22], [23]. This difficulty will be addressed by a more careful choice of the step-size policy; i.e, assumption 4.1, which is stronger than the more standard diminishing step-size scheme, e.g., in the DLPDS algorithm and [23]. We require this condition in order to prove, in the absence of the boundedness of $\{\mathcal{S}_x^{[i]}(k)\}$, the existence of a number of limits and summability of expansions toward Theorem 4.2. Thirdly, the DPPDS algorithm adopts the penalty relaxation instead of the Lagrangian relaxation in [22]. •

Remark 4.3: Observe that $\mu^{[i]}(k) \geq 0$, $\lambda^{[i]}(k) \geq 0$ and $v_x^{[i]}(k) \in X$ (due to the fact that X is convex). Furthermore, $([g(v_x^{[i]}(k))]^+, |h(v_x^{[i]}(k))|)$ is a supgradient of $\mathcal{H}^{[i]}_{v_x^{[i]}(k)}(w^{[i]}(k))$; i.e. the following *penalty supgradient inequality* holds for any $\mu \in \mathbb{R}^m_{\geq 0}$ and $\lambda \in \mathbb{R}^\nu_{\geq 0}$:

$$
\begin{aligned}
&([g(v_x^{[i]}(k))]^+)^T(\mu - v_\mu^{[i]}(k)) + |h(v_x^{[i]}(k))|^T(\lambda - v_\lambda^{[i]}(k)) \\
&\geq \mathcal{H}^{[i]}(v_x^{[i]}(k), \mu, \lambda) - \mathcal{H}^{[i]}(v_x^{[i]}(k), v_\mu^{[i]}(k), v_\lambda^{[i]}(k)). \qquad (11)
\end{aligned}
$$

•

Remark 4.4: A step-size sequence that satisfies the step-size assumption 4.1 is the harmonic series $\{\alpha(k) = \frac{1}{k+1}\}_{k\in\mathbb{Z}_{\geq 0}}$. It is obvious that $\lim\limits_{k\to+\infty} \dfrac{1}{k+1} = 0$, and well-known that $\sum_{k=0}^{+\infty} \frac{1}{k+1} = +\infty$ and $\sum_{k=0}^{+\infty} \frac{1}{(k+1)^2} < +\infty$. We now proceed to check the property of $\lim\limits_{k\to+\infty} \alpha(k+1)s(k) = 0$. For any $k \geq 1$, there is an integer $n \geq 1$ such that $2^{n-1} \leq k < 2^n$. It holds that

$$\begin{aligned}
&s(k) \leq s(2^n) = 1 + \frac{1}{2} + (\frac{1}{3} + \frac{1}{4}) + \cdots + (\frac{1}{2^{n-1}+1} + \cdots + \frac{1}{2^n})\\
&\leq 1 + \frac{1}{2} + (\frac{1}{3} + \frac{1}{3}) + \cdots + (\frac{1}{2^{n-1}+1} + \cdots + \frac{1}{2^{n-1}+1})\\
&\leq 1 + 1 + 1 + \cdots + 1 = n \leq \log_2 k + 1.
\end{aligned}$$

Then we have $\limsup\limits_{k\to+\infty} \dfrac{s(k)}{k+2} \leq \lim\limits_{k\to+\infty} \dfrac{\log_2 k + 1}{k+2} = 0$. Obviously, $\liminf\limits_{k\to+\infty} \dfrac{s(k)}{k+2} \geq 0$. Then we have the property of $\lim\limits_{k\to+\infty} \alpha(k+1)s(k) = 0$. Since $\log_2 k \leq (\log_2 k)^2 < (k+2)^{\frac{1}{2}}$, then

$$\begin{aligned}
&\sum_{k=0}^{+\infty} \alpha(k+1)^2 s(k)^2 \leq \sum_{k=0}^{+\infty} \frac{(\log_2 k + 1)^2}{(k+2)^2} = \sum_{k=0}^{+\infty} (\frac{(\log_2 k)^2}{(k+2)^2} + \frac{2\log_2 k}{(k+2)^2} + \frac{1}{(k+2)^2})\\
&\leq \sum_{k=0}^{+\infty} \frac{1}{(k+2)^{\frac{3}{2}}} + \sum_{k=0}^{+\infty} \frac{2}{(k+2)^{\frac{3}{2}}} + \sum_{k=0}^{+\infty} \frac{1}{(k+2)^2} < +\infty.
\end{aligned}$$

Additionally, we have $\sum_{k=0}^{+\infty} \alpha(k+1)^2 s(k) \leq \sum_{k=0}^{+\infty} \alpha(k+1)^2 s(k)^2 < +\infty$. •

V. Convergence analysis

In this sectiob, we provide the proofs for the main results, Theorem 3.2 and 4.2, of this paper. We start our analysis by providing some useful properties of the sequences weighted by $\{\alpha(k)\}$.

Lemma 5.1 (Convergence properties of weighted sequences): Let $K \geq 0$. Consider the sequence $\{\delta(k)\}$ defined by $\delta(k) := \frac{\sum_{\ell=K}^{k-1}\alpha(\ell)\rho(\ell)}{\sum_{\ell=K}^{k-1}\alpha(\ell)}$ where $k \geq K+1$, $\alpha(k) > 0$ and $\sum_{k=K}^{+\infty}\alpha(k) = +\infty$.

(a) If $\lim\limits_{k\to+\infty} \rho(k) = +\infty$, then $\lim\limits_{k\to+\infty} \delta(k) = +\infty$.

(b) If $\lim\limits_{k\to+\infty} \rho(k) = \rho^*$, then $\lim\limits_{k\to+\infty} \delta(k) = \rho^*$.

Proof: (a) For any $\Pi > 0$, there exists $k_1 \geq K$ such that $\rho(k) \geq \Pi$ for all $k \geq k_1$. Then the following holds for all $k \geq k_1 + 1$:

$$\delta(k) \geq \frac{1}{\sum_{\ell=K}^{k-1}\alpha(\ell)}(\sum_{\ell=K}^{k_1-1}\alpha(\ell)\rho(\ell) + \sum_{\ell=k_1}^{k-1}\alpha(\ell)\Pi) = \Pi + \frac{1}{\sum_{\ell=K}^{k-1}\alpha(\ell)}(\sum_{\ell=K}^{k_1-1}\alpha(\ell)\rho(\ell) - \sum_{\ell=K}^{k_1-1}\alpha(\ell)\Pi).$$

Take the limit on k in the above estimate and we have $\liminf_{k\to+\infty} \delta(k) \geq \Pi$. Since Π is arbitrary, then $\lim_{k\to+\infty} \delta(k) = +\infty$.

(b) For any $\epsilon > 0$, there exists $k_2 \geq K$ such that $\|\rho(k) - \rho^*\| \leq \epsilon$ for all $k \geq k_2 + 1$. Then we have

$$\|\delta(k) - \rho^*\| = \|\frac{\sum_{\tau=K}^{k-1} \alpha(\tau)(\rho(\tau) - \rho^*)}{\sum_{\tau=K}^{k-1} \alpha(\tau)}\|$$
$$\leq \frac{1}{\sum_{\tau=K}^{k-1} \alpha(\tau)}(\sum_{\tau=K}^{k_2-1} \alpha(\tau)\|\rho(\tau) - \rho^*\| + \sum_{\tau=k_2}^{k-1} \alpha(\tau)\epsilon) \leq \frac{\sum_{\tau=K}^{k_2-1} \alpha(\tau)\|\rho(\tau) - \rho^*\|}{\sum_{\tau=K}^{k-1} \alpha(\tau)} + \epsilon.$$

Take the limit on k in the above estimate and we have $\limsup_{k\to+\infty} \|\delta(k) - \rho^*\| \leq \epsilon$. Since ϵ is arbitrary, then $\lim_{k\to+\infty} \|\delta(k) - \rho^*\| = 0$. ∎

A. Proofs of Theorem 3.2

We now proceed to show Theorem 3.2. To do that, we first rewrite the DLPDS algorithm into the following form:

$$x^{[i]}(k+1) = v_x^{[i]}(k) + e_x^{[i]}(k), \quad \mu^{[i]}(k+1) = v_\mu^{[i]}(k) + e_\mu^{[i]}(k), \quad y^{[i]}(k+1) = v_y^{[i]}(k) + u^{[i]}(k),$$

where $e_x^{[i]}(k)$ and $e_\mu^{[i]}(k)$ are projection errors described by

$$e_x^{[i]}(k) := P_{X^{[i]}}[v_x^{[i]}(k) - \alpha(k)\mathcal{D}_x^{[i]}(k)] - v_x^{[i]}(k), \quad e_\mu^{[i]}(k) := P_{M^{[i]}}[v_\mu^{[i]}(k) + \alpha(k)\mathcal{D}_\mu^{[i]}(k)] - v_\mu^{[i]}(k),$$

and $u^{[i]}(k) := N(f^{[i]}(x^{[i]}(k)) - f^{[i]}(x^{[i]}(k-1)))$ is the local input which allows agent i to track the variation of the local objective function $f^{[i]}$. In this manner, the update law of each estimate is decomposed in two parts: a convex sum to fuse the information of each agent with those of its neighbors, plus some local error or input. With this decomposition, all the update laws are put into the same form as the dynamic average consensus algorithm in the Appendix. This observation allows us to divide the analysis of the DLPDS algorithm in two steps. Firstly, we show all the estimates asymptotically achieve consensus by utilizing the property that the local errors and inputs are diminishing. Secondly, we further show that the consensus vectors coincide with a pair of primal and Lagrangian dual optimal solutions and the optimal value.

Lemma 5.2 (Lipschitz continuity of $\mathcal{L}_x^{[i]}$ and $\mathcal{L}_\mu^{[i]}$): Consider $\mathcal{L}_\mu^{[i]}$ and $\mathcal{L}_x^{[i]}$. Then there are $L > 0$ and $R > 0$ such that $\|\mathcal{D}\mathcal{L}_\mu^{[i]}(x)\| \leq L$ and $\|\mathcal{D}\mathcal{L}_x^{[i]}(\mu)\| \leq R$ for each pair of $x \in \mathrm{co}(\cup_{i=1}^N X^{[i]})$ and $\mu \in \mathrm{co}(\cup_{i=1}^N M^{[i]})$. Furthermore, for each $\mu \in \mathrm{co}(\cup_{i=1}^N M^{[i]})$, the function $\mathcal{L}_\mu^{[i]}$ is

Lipschitz continuous with Lipschitz constant L over $\mathrm{co}(\cup_{i=1}^N X^{[i]})$, and for each $x \in \mathrm{co}(\cup_{i=1}^N X^{[i]})$, the function $\mathcal{L}_x^{[i]}$ is Lipschitz continuous with Lipschitz constant R over $\mathrm{co}(\cup_{i=1}^N M^{[i]})$.

Proof: Observe that $\mathcal{D}\mathcal{L}_\mu^{[i]} = \mathcal{D}f^{[i]} + \mu^T\mathcal{D}g$ and $\mathcal{D}\mathcal{L}_x^{[i]} = g$. Since $f^{[i]}$ and g_ℓ are convex, it follows from Proposition 5.4.2 in [3] that $\partial f^{[i]}$ and ∂g_ℓ are bounded over the compact $\mathrm{co}(\cup_{i=1}^N X^{[i]})$. Since $\mathrm{co}(\cup_{i=1}^N M^{[i]})$ is bounded, so is $\partial\mathcal{L}_\mu^{[i]}$, i.e., for any $\mu \in \mathrm{co}(\cup_{i=1}^N M^{[i]})$, there exists $L > 0$ such that $\|\partial\mathcal{L}_\mu^{[i]}(x)\| \leq L$ for all $x \in \mathrm{co}(\cup_{i=1}^N X^{[i]})$. Since g_ℓ is continuous (due to its convexity) and $\mathrm{co}(\cup_{i=1}^N X^{[i]})$ is bounded, then g and thus $\partial\mathcal{L}_x^{[i]}$ are bounded, i.e., for any $x \in \mathrm{co}(\cup_{i=1}^N X^{[i]})$, there exists $R > 0$ such that $\|\partial\mathcal{L}_x^{[i]}(\mu)\| \leq R$ for all $\mu \in \mathrm{co}(\cup_{i=1}^N M^{[i]})$.

It follows from the Lagrangian subgradient inequality that

$$\mathcal{D}\mathcal{L}_\mu^{[i]}(x)^T(x' - x) \leq \mathcal{L}_\mu^{[i]}(x') - \mathcal{L}_\mu^{[i]}(x), \quad \mathcal{D}\mathcal{L}_\mu^{[i]}(x')^T(x - x') \leq \mathcal{L}_\mu^{[i]}(x) - \mathcal{L}_\mu^{[i]}(x'),$$

for any $x, x' \in \mathrm{co}(\cup_{i=1}^N X^{[i]})$. By using the boundedness of the subdifferentials, the above two inequalities give that $-L\|x - x'\| \leq \mathcal{L}_\mu^{[i]}(x) - \mathcal{L}_\mu^{[i]}(x') \leq L\|x - x'\|$. This implies that $\|\mathcal{L}_\mu^{[i]}(x) - \mathcal{L}_\mu^{[i]}(x')\| \leq L\|x - x'\|$ for any $x, x' \in \mathrm{co}(\cup_{i=1}^m X^{[i]})$. The proof for the Lipschitz continuity of $\mathcal{L}_x^{[i]}$ is analogous by using the Lagrangian supgradient inequality. ■

The following lemma provides a basic iteration relation used in the convergence proof for the DLPDS algorithm.

Lemma 5.3 (Basic iteration relation): Let the balanced communication assumption 2.3 and the periodic strong connectivity assumption 2.4 hold. For any $x \in X$, any $\mu \in M$ and all $k \geq 0$, the following estimates hold:

$$\begin{aligned}\sum_{i=1}^N \|e_x^{[i]}(k) + \alpha(k)\mathcal{D}_x^{[i]}(k)\|^2 \leq{}& \sum_{i=1}^N \alpha(k)^2\|\mathcal{D}_x^{[i]}(k)\|^2 + \sum_{i=1}^N \{\|x^{[i]}(k) - x\|^2 - \|x^{[i]}(k+1) - x\|^2\} \\ & - \sum_{i=1}^N 2\alpha(k)(\mathcal{L}^{[i]}(v_x^{[i]}(k), v_\mu^{[i]}(k)) - \mathcal{L}^{[i]}(x, v_\mu^{[i]}(k))),\end{aligned} \tag{12}$$

$$\begin{aligned}\sum_{i=1}^N \|e_\mu^{[i]}(k) - \alpha(k)\mathcal{D}_\mu^{[i]}(k)\|^2 \leq{}& \sum_{i=1}^N \alpha(k)^2\|\mathcal{D}_\mu^{[i]}(k)\|^2 + \sum_{i=1}^N \{\|\mu^{[i]}(k) - \mu\|^2 - \|\mu^{[i]}(k+1) - \mu\|^2\} \\ & + \sum_{i=1}^N 2\alpha(k)(\mathcal{L}^{[i]}(v_x^{[i]}(k), v_\mu^{[i]}(k)) - \mathcal{L}^{[i]}(v_x^{[i]}(k), \mu)).\end{aligned} \tag{13}$$

Proof: By Lemma 9.1 with $Z = M^{[i]}$, $z = v_\mu^{[i]}(k) + \alpha(k)\mathcal{D}_\mu^{[i]}(k)$ and $y = \mu \in M$, we have that for all $k \geq 0$

$$\sum_{i=1}^{N} \|e_\mu^{[i]}(k) - \alpha(k)\mathcal{D}_\mu^{[i]}(k)\|^2 \leq \sum_{i=1}^{N} \|v_\mu^{[i]}(k) + \alpha(k)\mathcal{D}_\mu^{[i]}(k) - \mu\|^2 - \sum_{i=1}^{N} \|\mu^{[i]}(k+1) - \mu\|^2$$
$$= \sum_{i=1}^{N} \|v_\mu^{[i]}(k) - \mu\|^2 + \sum_{i=1}^{N} \alpha(k)^2 \|\mathcal{D}_\mu^{[i]}(k)\|^2$$
$$+ \sum_{i=1}^{N} 2\alpha(k)\mathcal{D}_\mu^{[i]}(k)^T (v_\mu^{[i]}(k) - \mu) - \sum_{i=1}^{N} \|\mu^{[i]}(k+1) - \mu\|^2$$
$$\leq \sum_{i=1}^{N} \alpha(k)^2 \|\mathcal{D}_\mu^{[i]}(k)\|^2 + \sum_{i=1}^{N} 2\alpha(k)\mathcal{D}_\mu^{[i]}(k)^T (v_\mu^{[i]}(k) - \mu)$$
$$+ \sum_{i=1}^{N} \|\mu^{[i]}(k) - \mu\|^2 - \sum_{i=1}^{N} \|\mu^{[i]}(k+1) - \mu\|^2. \tag{14}$$

One can show (13) by substituting the following Lagrangian supgradient inequality into (14):

$$\mathcal{D}_\mu^{[i]}(k)^T (\mu - v_\mu^{[i]}(k)) \geq \mathcal{L}^{[i]}(v_x^{[i]}(k), \mu) - \mathcal{L}^{[i]}(v_x^{[i]}(k), v_\mu^{[i]}(k)).$$

Similarly, equality (12) can be shown by using the following Lagrangian subgradient inequality: $\mathcal{D}_x^{[i]}(k)^T (x - v_x^{[i]}(k)) \leq \mathcal{L}^{[i]}(x, v_\mu^{[i]}(k)) - \mathcal{L}^{[i]}(v_x^{[i]}(k), v_\mu^{[i]}(k))$. ∎

The following lemma shows that the consensus is asymptotically reached.

Lemma 5.4 (Achieving consensus): Let the non-degeneracy assumption 2.2, the balanced communication assumption 2.3 and the periodic strong connectivity assumption 2.4 hold. Consider the sequences of $\{x^{[i]}(k)\}$, $\{\mu^{[i]}(k)\}$ and $\{y^{[i]}(k)\}$ of the DLPDS algorithm with the step-size sequence $\{\alpha(k)\}$ satisfying $\lim_{k\to+\infty} \alpha(k) = 0$. Then there exist $x^* \in X$ and $\mu^* \in M$ such that $\lim_{k\to+\infty} \|x^{[i]}(k) - x^*\| = 0$, $\lim_{k\to+\infty} \|\mu^{[i]}(k) - \mu^*\| = 0$ for all $i \in V$, and $\lim_{k\to+\infty} \|y^{[i]}(k) - y^{[j]}(k)\| = 0$ for all $i, j \in V$.

Proof: Observe that $v_x^{[i]}(k) \in \mathrm{co}(\cup_{i=1}^N X^{[i]})$ and $v_\mu^{[i]}(k) \in \mathrm{co}(\cup_{i=1}^N M^{[i]})$. Then it follows from Lemma 5.2 that $\|\mathcal{D}_x^{[i]}(k)\| \leq L$. From Lemma 5.3 it follows that

$$\sum_{i=1}^{N} \|x^{[i]}(k+1) - x\|^2 \leq \sum_{i=1}^{N} \|x^{[i]}(k) - x\|^2 + \sum_{i=1}^{N} \alpha(k)^2 L^2$$
$$+ \sum_{i=1}^{N} 2\alpha(k)(\|\mathcal{L}^{[i]}(v_x^{[i]}(k), v_\mu^{[i]}(k))\| + \|\mathcal{L}^{[i]}(x, v_\mu^{[i]}(k))\|). \tag{15}$$

Notice that $v_x^{[i]}(k) \in \mathrm{co}(\cup_{i=1}^N X^{[i]})$, $v_\mu^{[i]}(k) \in \mathrm{co}(\cup_{i=1}^N M^{[i]})$ and $x \in X$ are bounded. Since $\mathcal{L}^{[i]}$ is continuous, then $\mathcal{L}^{[i]}(v_x^{[i]}(k), v_\mu^{[i]}(k))$ and $\mathcal{L}^{[i]}(x, v_\mu^{[i]}(k))$ are bounded. Since $\lim_{k\to+\infty} \alpha(k) = 0$, the last two terms on the right-hand side of (15) converge to zero as $k \to +\infty$. Taking limits on both sides of (15), one can see that $\limsup_{k\to+\infty} \sum_{i=1}^N \|x^{[i]}(k+1) - x\|^2 \leq \liminf_{k\to+\infty} \sum_{i=1}^N \|x^{[i]}(k) - x\|^2$ for any $x \in X$, and thus $\lim_{k\to+\infty} \sum_{i=1}^N \|x^{[i]}(k) - x\|^2$ exists for any $x \in X$. On the other hand, taking limits on both sides of (12) we obtain $\lim_{k\to+\infty} \sum_{i=1}^N \|e_x^{[i]}(k) + \alpha(k)\mathcal{D}_x^{[i]}(k)\|^2 = 0$ and therefore we deduce that $\lim_{k\to+\infty} \|e_x^{[i]}(k)\| = 0$ for all $i \in V$. It follows from Proposition 9.1 in the Appendix that $\lim_{k\to+\infty} \|x^{[i]}(k) - x^{[j]}(k)\| = 0$ for all $i, j \in V$. Combining this with the property that $\lim_{k\to+\infty} \|x^{[i]}(k) - x\|$ exists for any $x \in X$, we deduce that there exists $x^* \in \mathbb{R}^n$ such that $\lim_{k\to+\infty} \|x^{[i]}(k) - x^*\| = 0$ for all $i \in V$. Since $x^{[i]}(k) \in X^{[i]}$ and $X^{[i]}$ is closed, it implies that $x^* \in X^{[i]}$ for all $i \in V$ and thus $x^* \in X$. Similarly, one can show that there is $\mu^* \in M$ such that $\lim_{k\to+\infty} \|\mu^{[i]}(k) - \mu^*\| = 0$ for all $i \in V$.

Since $\lim_{k\to+\infty} \|x^{[i]}(k) - x^*\| = 0$ and $f^{[i]}$ is continuous, then $\lim_{k\to+\infty} \|u^{[i]}(k)\| = 0$. It follows from Proposition 9.1 that $\lim_{k\to+\infty} \|y^{[i]}(k) - y^{[j]}(k)\| = 0$ for all $i, j \in V$. ■

From Lemma 5.4, we know that the sequences of $\{x^{[i]}(k)\}$ and $\{\mu^{[i]}(k)\}$ of the DLPDS algorithm asymptotically agree on to some point in X and some point in M, respectively. Denote by $\Theta \subseteq X \times M$ the set of such limit points. We further denote by the average of primal and dual estimates $\hat{x}(k) := \frac{1}{N}\sum_{i=1}^N x^{[i]}(k)$ and $\hat{\mu}(k) := \frac{1}{N}\sum_{i=1}^N \mu^{[i]}(k)$, respectively. The following lemma further characterizes that the points in Θ are saddle points of the Lagrangian function $\mathcal{L}$ over $X \times M$.

Lemma 5.5 (Saddle-point characterization of Θ): Each point in Θ is a saddle point of the Lagrangian function $\mathcal{L}$ over $X \times M$.

Proof: Denote by the maximum deviation of primal estimates $\Delta_x(k) := \max_{i,j\in V} \|x^{[j]}(k) - x^{[i]}(k)\|$. Notice that

$$\begin{aligned}
&\|v_x^{[i]}(k) - \hat{x}(k)\| = \|\sum_{j=1}^{N} a_j^i(k)x^{[j]}(k) - \sum_{j=1}^{N}\frac{1}{N}x^{[j]}(k)\| \\
&= \|\sum_{j\neq i} a_j^i(k)(x^{[j]}(k) - x^{[i]}(k)) - \sum_{j\neq i}\frac{1}{N}(x^{[j]}(k) - x^{[i]}(k))\| \\
&\leq \sum_{j\neq i} a_j^i(k)\|x^{[j]}(k) - x^{[i]}(k)\| + \sum_{j\neq i}\frac{1}{N}\|x^{[j]}(k) - x^{[i]}(k)\| \leq 2\Delta_x(k).
\end{aligned}$$

Denote by the maximum deviation of dual estimates $\Delta_\mu(k) := \max_{i,j\in V} \|\mu^{[j]}(k) - \mu^{[i]}(k)\|$. Similarly, we have $\|v_\mu^{[i]}(k) - \hat{\mu}(k)\| \leq 2\Delta_\mu(k)$.

We will show this lemma by contradiction. Suppose that there is $(x^*, \mu^*) \in \Theta$ which is not a saddle point of $\mathcal{L}$ over $X \times M$. Then at least one of the following equalities holds:

$$\exists x \in X \quad \text{s.t.} \quad \mathcal{L}(x^*, \mu^*) > \mathcal{L}(x, \mu^*), \tag{16}$$

$$\exists \mu \in M \quad \text{s.t.} \quad \mathcal{L}(x^*, \mu) > \mathcal{L}(x^*, \mu^*). \tag{17}$$

Suppose first that (16) holds. Then, there exists $\varsigma > 0$ such that $\mathcal{L}(x^*, \mu^*) = \mathcal{L}(x, \mu^*) + \varsigma$. Consider the sequences of $\{x^{[i]}(k)\}$ and $\{\mu^{[i]}(k)\}$ which converge respectively to x^* and μ^* defined above. Notice that estimate (12) leads to

$$\begin{aligned}
&\sum_{i=1}^{N} \|x^{[i]}(k+1) - x\|^2 \leq \sum_{i=1}^{N} \|x^{[i]}(k) - x\|^2 \; + \alpha(k)^2 \sum_{i=1}^{N} \|\mathcal{D}_x^{[i]}(k)\|^2 \\
&- 2\alpha(k)\sum_{i=1}^{N}(A_i(k) + B_i(k) + C_i(k) + D_i(k) + E_i(k) + F_i(k)),
\end{aligned}$$

where

$$A_i(k) := \mathcal{L}^{[i]}(v_x^{[i]}(k), v_\mu^{[i]}(k)) - \mathcal{L}^{[i]}(\hat{x}(k), v_\mu^{[i]}(k)), \;\; B_i(k) := \mathcal{L}^{[i]}(\hat{x}(k), v_\mu^{[i]}(k)) - \mathcal{L}^{[i]}(\hat{x}(k), \hat{\mu}(k)),$$

$$C_i(k) := \mathcal{L}^{[i]}(\hat{x}(k), \hat{\mu}(k)) - \mathcal{L}^{[i]}(x^*, \hat{\mu}(k)), \;\; D_i(k) := \mathcal{L}^{[i]}(x^*, \hat{\mu}(k)) - \mathcal{L}^{[i]}(x^*, \mu^*),$$

$$E_i(k) := \mathcal{L}^{[i]}(x^*, \mu^*) - \mathcal{L}^{[i]}(x, \mu^*), \;\; F_i(k) = \mathcal{L}^{[i]}(x, \mu^*) - \mathcal{L}^{[i]}(x, v_\mu^{[i]}(k)).$$

It follows from the Lipschitz continuity property of $\mathcal{L}^{[i]}$; see Lemma 5.2, that

$$\|A_i(k)\| \leq L\|v_x^{[i]}(k) - \hat{x}(k)\| \leq 2L\Delta_x(k), \quad \|B_i(k)\| \leq R\|v_\mu^{[i]}(k) - \hat{\mu}(k)\| \leq 2R\Delta_\mu(k),$$

$$\|C_i(k)\| \leq L\|\hat{x}(k) - x^*\| \leq \frac{L}{N}\sum_{i=1}^{N}\|x^{[i]}(k) - x^*\|,$$

$$\|D_i(k)\| \leq R\|\hat{\mu}(k) - \mu^*\| \leq \frac{R}{N}\sum_{i=1}^{N}\|\mu^{[i]}(k) - \mu^*\|,$$

$$\begin{aligned}\|F_i(k)\| &\leq R\|\mu^* - v_\mu^{[i]}(k)\| \leq R\|\mu^* - \hat{\mu}(k)\| + R\|\hat{\mu}(k) - v_\mu^{[i]}(k)\| \\ &\leq \frac{R}{N}\sum_{i=1}^{N}\|\mu^*(k) - \mu^{[i]}(k)\| + 2R\Delta_\mu(k).\end{aligned}$$

Since $\lim_{k\to+\infty}\|x^{[i]}(k) - x^*\| = 0$, $\lim_{k\to+\infty}\|\mu^{[i]}(k) - \mu^*\| = 0$, $\lim_{k\to+\infty}\Delta_x(k) = 0$ and $\lim_{k\to+\infty}\Delta_\mu(k) = 0$, then all $A_i(k), B_i(k), C_i(k), D_i(k), F_i(k)$ converge to zero as $k \to +\infty$. Then there exists $k_0 \geq 0$ such that for all $k \geq k_0$, it holds that

$$\sum_{i=1}^{N}\|x^{[i]}(k+1) - x\|^2 \leq \sum_{i=1}^{N}\|x^{[i]}(k) - x\|^2 + N\alpha(k)^2L^2 - \varsigma\alpha(k).$$

Following a recursive argument, we have that for all $k \geq k_0$, it holds that

$$\sum_{i=1}^{N}\|x^{[i]}(k+1) - x\|^2 \leq \sum_{i=1}^{N}\|x^{[i]}(k_0) - x\|^2 + NL^2\sum_{\tau=k_0}^{k}\alpha(\tau)^2 - \varsigma\sum_{\tau=k_0}^{k}\alpha(\tau).$$

Since $\sum_{k=k_0}^{+\infty}\alpha(k) = +\infty$ and $\sum_{k=k_0}^{+\infty}\alpha(k)^2 < +\infty$ and $x^{[i]}(k_0) \in X^{[i]}$, $x \in X$ are bounded, the above estimate yields a contradiction by taking k sufficiently large. In other words, (16) cannot hold. Following a parallel argument, one can show that (17) cannot hold either. This ensures that each $(x^*, \mu^*) \in \Theta$ is a saddle point of $\mathcal{L}$ over $X \times M$. ■

The combination of (c) in Lemmas 3.1 and Lemma 5.5 gives that, for each $(x^*, \mu^*) \in \Theta$, we have that $\mathcal{L}(x^*, \mu^*) = p^*$ and μ^* is Lagrangian dual optimal. We still need to verify that x^* is a primal optimal solution. We are now in the position to show Theorem 3.2 based on the following two claims.

Proofs of Theorem 3.2:

Claim 1: Each point $(x^*, \mu^*) \in \Theta$ is a point in $X^* \times D_L^*$.

Proof: The Lagrangian dual optimality of μ^* follows from (c) in Lemma 3.1 and Lemma 5.5. To characterize the primal optimality of x^*, we define an auxiliary sequence $\{z(k)\}$ by $z(k) :=$

$\frac{\sum_{\tau=0}^{k-1}\alpha(\tau)\hat{x}(\tau)}{\sum_{\tau=0}^{k-1}\alpha(\tau)}$ which is a weighted version of the average of primal estimates. Since $\lim_{k\to+\infty}\hat{x}(k)=x^*$, it follows from Lemma 5.1 (b) that $\lim_{k\to+\infty}z(k)=x^*$.

Since (x^*,μ^*) is a saddle point of $\mathcal{L}$ over $X\times M$, then $\mathcal{L}(x^*,\mu)\leq\mathcal{L}(x^*,\mu^*)$ for any $\mu\in M$; i.e., the following relation holds for any $\mu\in M$:

$$g(x^*)^T(\mu-\mu^*)\leq 0. \tag{18}$$

Choose $\mu_a=\mu^*+\min_{i\in V}\theta^{[i]}\frac{\mu^*}{\|\mu^*\|}$ where $\theta^{[i]}>0$ is given in the definition of $M^{[i]}$. Then $\mu_a\geq 0$ and $\|\mu_a\|\leq\|\mu^*\|+\min_{i\in V}\theta^{[i]}$ implying $\mu_a\in M$. Letting $\mu=\mu_a$ in (18) gives that

$$\frac{\min_{i\in V}\theta^{[i]}}{\|\mu^*\|}g(x^*)^T\mu^*\leq 0.$$

Since $\theta^{[i]}>0$, we have $g(x^*)^T\mu^*\leq 0$. On the other hand, we choose $\mu_b=\frac{1}{2}\mu^*$ and then $\mu_b\in M$. Letting $\mu=\mu_b$ in (18) gives that $-\frac{1}{2}g(x^*)^T\mu^*\leq 0$ and thus $g(x^*)^T\mu^*\geq 0$. The combination of the above two estimates guarantees the property of $g(x^*)^T\mu^*=0$.

We now proceed to show $g(x^*)\leq 0$ by contradiction. Assume that $g(x^*)\leq 0$ does not hold. Denote $J^+(x^*):=\{1\leq\ell\leq m\mid g_\ell(x^*)>0\}\neq\emptyset$ and $\eta:=\min_{\ell\in J^+(x^*)}\{g_\ell(x^*)\}$. Then $\eta>0$. Since g is continuous and $v_x^{[i]}(k)$ converges to x^*, there exists $K\geq 0$ such that $g_\ell(v_x^{[i]}(k))\geq\frac{\eta}{2}$ for all $k\geq K$ and all $\ell\in J^+(x^*)$. Since $v_\mu^{[i]}(k)$ converges to μ^*, without loss of generality, we say that $\|v_\mu^{[i]}(k)-\mu^*\|\leq\frac{1}{2}\min_{i\in V}\theta^{[i]}$ for all $k\geq K$. Choose $\hat{\mu}$ such that $\hat{\mu}_\ell=\mu_\ell^*$ for $\ell\notin J^+(x^*)$ and $\hat{\mu}_\ell=\mu_\ell^*+\frac{1}{\sqrt{m}}\min_{i\in V}\theta^{[i]}$ for $\ell\in J^+(x^*)$. Since $\mu^*\geq 0$ and $\theta^{[i]}>0$, thus $\hat{\mu}\geq 0$. Furthermore, $\|\hat{\mu}\|\leq\|\mu^*\|+\min_{i\in V}\theta^{[i]}$, then $\hat{\mu}\in M$. Equating μ to $\hat{\mu}$ and letting $\mathcal{D}_\mu^{[i]}(k)=g(v_x^{[i]}(k))$ in the estimate (14), the following holds for $k\geq K$:

$$\begin{aligned}
N|J^+(x^*)|\eta\min_{i\in V}\theta^{[i]}\alpha(k)&\leq 2\alpha(k)\sum_{i=1}^N\sum_{\ell\in J^+(x^*)}g_\ell(v_x^{[i]}(k))(\hat{\mu}-v_\mu^{[i]}(k))_\ell\\
&\leq\sum_{i=1}^N\|\mu^{[i]}(k)-\hat{\mu}\|^2-\sum_{i=1}^N\|\mu^{[i]}(k+1)-\hat{\mu}\|^2+NR^2\alpha(k)^2\\
&-2\alpha(k)\sum_{i=1}^N\sum_{\ell\notin J^+(x^*)}g_\ell(v_x^{[i]}(k))(\hat{\mu}-v_\mu^{[i]}(k))_\ell.
\end{aligned}\tag{19}$$

Summing (19) over $[K, k-1]$ with $k \geq K+1$, dividing by $\sum_{\tau=K}^{k-1} \alpha(\tau)$ on both sides, and using $-\sum_{i=1}^{N} \|\mu^{[i]}(k) - \hat{\mu}\|^2 \leq 0$, we obtain

$$N|J^+(x^*)|\eta \min_{i\in V} \theta^{[i]} \leq \frac{1}{\sum_{\tau=K}^{k-1} \alpha(\tau)} \{ \sum_{i=1}^{N} \|\mu^{[i]}(K) - \hat{\mu}\|^2 + NR^2 \sum_{\tau=K}^{k-1} \alpha(\tau)^2$$
$$- \sum_{\tau=K}^{k-1} 2\alpha(\tau) \sum_{i=1}^{N} \sum_{\ell \notin J^+(x^*)} g_\ell(v_x^{[i]}(\tau))(\hat{\mu} - v_\mu^{[i]}(\tau))_\ell \}. \quad (20)$$

Since $\mu^{[i]}(K) \in M^{[i]}$, $\hat{\mu} \in M$ are bounded and $\sum_{\tau=K}^{+\infty} \alpha(\tau) = +\infty$, then the limit of the first term on the right hand side of (20) is zero as $k \to +\infty$. Since $\sum_{\tau=K}^{+\infty} \alpha(\tau)^2 < +\infty$, then the limit of the second term is zero as $k \to +\infty$. Since $\lim_{k\to+\infty} v_x^{[i]}(k) = x^*$ and $\lim_{k\to+\infty} v_\mu^{[i]}(k) = \mu^*$, thus $\lim_{k\to+\infty} 2 \sum_{i=1}^{N} \sum_{\ell \notin J^+(x^*)} g_\ell(v_x^{[i]}(k))(\hat{\mu} - v_\mu^{[i]}(k))_\ell = 0$. Then it follows from Lemma 5.1 (b) that then the limit of the third term is zero as $k \to +\infty$. Then we have $N|J^+(x^*)|\eta \min_{i\in V} \theta^{[i]} \leq 0$. Recall that $|J^+(x^*)| > 0$, $\eta > 0$ and $\theta^{[i]} > 0$. Then we reach a contradiction, implying that $g(x^*) \leq 0$.

Since $x^* \in X$ and $g(x^*) \leq 0$, then x^* is a feasible solution and thus $f(x^*) \geq p^*$. On the other hand, since $z(k)$ is a convex combination of $\hat{x}(0), \cdots, \hat{x}(k-1)$ and f is convex, thus we have the following estimate:

$$f(z(k)) \leq \frac{\sum_{\tau=0}^{k-1} \alpha(\tau) f(\hat{x}(\tau))}{\sum_{\tau=0}^{k-1} \alpha(\tau)} = \frac{1}{\sum_{\tau=0}^{k-1} \alpha(\tau)} \{ \sum_{\tau=0}^{k-1} \alpha(\tau) \mathcal{L}(\hat{x}(\tau), \hat{\mu}(\tau)) - \sum_{\tau=0}^{k-1} N\alpha(\tau) \hat{\mu}(\tau)^T g(\hat{x}(\tau)) \}.$$

Recall the following convergence properties:

$$\lim_{k\to+\infty} z(k) = x^*, \quad \lim_{k\to+\infty} \mathcal{L}(\hat{x}(k), \hat{\mu}(k)) = \mathcal{L}(x^*, \mu^*) = p^*, \quad \lim_{k\to+\infty} \hat{\mu}(k)^T g(\hat{x}(k)) = g(x^*)^T \mu^* = 0.$$

It follows from Lemma 5.1 (b) that $f(x^*) \leq p^*$. Therefore, we have $f(x^*) = p^*$, and thus x^* is a primal optimal point. ■

Claim 2: It holds that $\lim_{k\to+\infty} \|y^{[i]}(k) - p^*\| = 0$.

Proof: The following can be proven by induction on k for a fixed $k' \geq 1$:

$$\sum_{i=1}^{N} y^{[i]}(k+1) = \sum_{i=1}^{N} y^{[i]}(k') + N \sum_{\ell=k'}^{k} \sum_{i=1}^{N} (f^{[i]}(x^{[i]}(\ell)) - f^{[i]}(x^{[i]}(\ell-1))). \quad (21)$$

Let $k' = 1$ in (21) and recall that initial state $y^{[i]}(1) = Nf^{[i]}(x^{[i]}(0))$ for all $i \in V$. Then we have

$$\sum_{i=1}^{N} y^{[i]}(k+1) = \sum_{i=1}^{N} y^{[i]}(1) + N \sum_{i=1}^{N} (f^{[i]}(x^{[i]}(k)) - f^{[i]}(x^{[i]}(0))) = N \sum_{i=1}^{N} f^{[i]}(x^{[i]}(k)). \quad (22)$$

The combination of (22) with $\lim_{k\to+\infty}\|y^{[i]}(k)-y^{[j]}(k)\|=0$ gives the desired result. We then finish the proofs of Theorem 3.2. ■

B. Proofs of Theorem 4.2

In this part, we present the proofs of Theorem 4.2. In order to analyze the DPPDS algorithm, we first rewrite it into the following form:

$$\mu^{[i]}(k+1)=v_\mu^{[i]}(k)+u_\mu^{[i]}(k),\quad \lambda^{[i]}(k+1)=v_\lambda^{[i]}(k)+u_\lambda^{[i]}(k),$$
$$x^{[i]}(k+1)=v_x^{[i]}(k)+e_x^{[i]}(k),\quad y^{[i]}(k+1)=v_y^{[i]}(k)+u_y^{[i]}(k),$$

where $e_x^{[i]}(k)$ is projection error described by

$$e_x^{[i]}(k):=P_X[v_x^{[i]}(k)-\alpha(k)\mathcal{S}_x^{[i]}(k)]-v_x^{[i]}(k),$$

and $u_\mu^{[i]}(k):=\alpha(k)[g(v_x^{[i]}(k))]^+$, $u_\lambda^{[i]}(k):=\alpha(k)|h(v_x^{[i]}(k))|$, $u_y^{[i]}(k)=N(f^{[i]}(x^{[i]}(k))-f^{[i]}(x^{[i]}(k-1)))$ are some local inputs. Denote by the maximum deviations of dual estimates $M_\mu(k):=\max_{i\in V}\|\mu^{[i]}(k)\|$ and $M_\lambda(k):=\max_{i\in V}\|\lambda^{[i]}(k)\|$. We further denote by the averages of primal and dual estimates $\hat{x}(k):=\frac{1}{N}\sum_{i=1}^N x^{[i]}(k)$, $\hat{\mu}(k):=\frac{1}{N}\sum_{i=1}^N\mu^{[i]}(k)$ and $\hat{\lambda}(k):=\frac{1}{N}\sum_{i=1}^N\lambda^{[i]}(k)$.

Before showing Lemma 5.6, we present some useful facts. Since X is compact, and $f^{[i]}$, $[g(\cdot)]^+$ and h are continuous, there exist $F,G^+,H>0$ such that for all $x\in X$, it holds that $\|f^{[i]}(x)\|\le F$ for all $i\in V$, $\|[g(x)]^+\|\le G^+$ and $\|h(x)\|\le H$. Since X is a compact set and $f^{[i]}$, $[g_\ell(\cdot)]^+$, $|h_\ell(\cdot)|$ are convex, then it follows from Proposition 5.4.2 in [3] that there exist $D_F,D_{G^+},D_H>0$ such that for all $x\in X$, it holds that $\|\mathcal{D}f^{[i]}(x)\|\le D_F$ ($i\in V$), $m\|\mathcal{D}[g_\ell(x)]^+\|\le D_{G^+}$ ($1\le\ell\le m$) and $\nu\|\mathcal{D}|h_\ell|(x)\|\le D_H$ ($1\le\ell\le\nu$).

Lemma 5.6 (Diminishing and summable properties): Suppose the balanced communication assumption 2.3 and the step-size assumption 4.1 hold.

(a) It holds that $\lim_{k\to+\infty}\alpha(k)M_\mu(k)=0$, $\lim_{k\to+\infty}\alpha(k)M_\lambda(k)=0$, $\lim_{k\to+\infty}\alpha(k)\|\mathcal{S}_x^{[i]}(k)\|=0$, and the sequences of $\{\alpha(k)^2M_\mu^2(k)\}$, $\{\alpha(k)^2M_\lambda^2(k)\}$ and $\{\alpha(k)^2\|\mathcal{S}_x^{[i]}(k)\|^2\}$ are summable.

(b) The sequences $\{\alpha(k)\|\hat{\mu}(k)-v_\mu^{[i]}(k)\|\}$, $\{\alpha(k)\|\hat{\lambda}(k)-v_\lambda^{[i]}(k)\|\}$, $\{\alpha(k)M_\mu(k)\|\hat{x}(k)-v_x^{[i]}(k)\|\}$, $\{\alpha(k)M_\lambda(k)\|\hat{x}(k)-v_x^{[i]}(k)\|\}$ and $\{\alpha(k)\|\hat{x}(k)-v_x^{[i]}(k)\|\}$ are summable.

Proof: (a) Notice that

$$\|v_\mu^{[i]}(k)\|=\|\sum_{j=1}^N a_j^i(k)\mu^{[j]}(k)\|\le\sum_{j=1}^N a_j^i(k)\|\mu^{[j]}(k)\|\le\sum_{j=1}^N a_j^i(k)M_\mu(k)=M_\mu(k),$$

where in the last equality we use the balanced communication assumption 2.3. Recall that $v_x^{[i]}(k) \in X$. This implies that the following inequalities hold for all $k \geq 0$:

$$\|\mu^{[i]}(k+1)\| \leq \|v_\mu^{[i]}(k) + \alpha(k)[g(v_x^{[i]}(k))]^+\| \leq \|v_\mu^{[i]}(k)\| + G^+\alpha(k) \leq M_\mu(k) + G^+\alpha(k).$$

From here, then we deduce the following recursive estimate on $M_\mu(k+1)$: $M_\mu(k+1) \leq M_\mu(k) + G^+\alpha(k)$. Repeatedly applying the above estimates yields that

$$M_\mu(k+1) \leq M_\mu(0) + G^+ s(k). \tag{23}$$

Similar arguments can be employed to show that

$$M_\lambda(k+1) \leq M_\lambda(0) + Hs(k). \tag{24}$$

Since $\lim_{k\to+\infty} \alpha(k+1)s(k) = 0$ and $\lim_{k\to+\infty} \alpha(k) = 0$, then we know that $\lim_{k\to+\infty} \alpha(k+1)M_\mu(k+1) = 0$ and $\lim_{k\to+\infty} \alpha(k+1)M_\lambda(k+1) = 0$. Notice that the following estimate on $\mathcal{S}_x^{[i]}(k)$ holds:

$$\|\mathcal{S}_x^{[i]}(k)\| \leq D_F + D_{G^+}M_\mu(k) + D_H M_\lambda(k). \tag{25}$$

Recall that $\lim_{k\to+\infty} \alpha(k) = 0$, $\lim_{k\to+\infty} \alpha(k)M_\mu(k) = 0$ and $\lim_{k\to+\infty} \alpha(k)M_\lambda(k) = 0$. Then the result of $\lim_{k\to+\infty} \alpha(k)\|\mathcal{S}_x^{[i]}(k)\| = 0$ follows. By (23), we have

$$\sum_{k=0}^{+\infty} \alpha(k)^2 M_\mu^2(k) \leq \alpha(0)^2 M_\mu^2(0) + \sum_{k=1}^{+\infty} \alpha(k)^2(M_\mu(0) + G^+ s(k-1))^2.$$

It follows from the step-size assumption 4.1 that $\sum_{k=0}^{+\infty} \alpha(k)^2 M_\mu^2(k) < +\infty$. Similarly, one can show that $\sum_{k=0}^{+\infty} \alpha(k)^2 M_\lambda^2(k) < +\infty$. By using (23), (24) and (25), we have the following estimate:

$$\sum_{k=0}^{+\infty} \alpha(k)^2 \|\mathcal{S}_x^{[i]}(k)\|^2 \leq \alpha(0)^2(D_F + D_{G^+}M_\mu(0) + D_H M_\lambda(0))^2$$
$$+ \sum_{k=1}^{+\infty} \alpha(k)^2(D_F + D_{G^+}(M_\mu(0) + G^+ s(k-1)) + D_H(M_\lambda(0) + Hs(k-1)))^2.$$

Then the summability of $\{\alpha(k)^2\}$, $\{\alpha(k+1)^2 s(k)\}$ and $\{\alpha(k+1)^2 s(k)^2\}$ verifies that of $\{\alpha(k)^2\|\mathcal{S}_x^{[i]}(k)\|^2\}$.

(b) Consider the dynamics of $\mu^{[i]}(k)$ which is in the same form as the distributed projected subgradient algorithm in [23]. Recall that $\{[g(v_x^{[i]}(k))]^+\}$ is uniformly bounded. Then following from Lemma 9.2 in the Appendix with $Z = \mathbb{R}^m_{\geq 0}$ and $d^{[i]}(k) = -[g(v_x^{[i]}(k))]^+$, we have the

summability of $\{\alpha(k)\max_{i\in V}\|\hat{\mu}(k)-\mu^{[i]}(k)\|\}$. Then $\{\alpha(k)\|\hat{\mu}(k)-v_{\mu}^{[i]}(k)\|\}$ is summable by using the following set of inequalities:

$$\|\hat{\mu}(k)-v_{\mu}^{[i]}(k)\| \leq \sum_{j=1}^{N} a_j^i(k)\|\hat{\mu}(k)-\mu^{[j]}(k)\| \leq \max_{i\in V}\|\hat{\mu}(k)-\mu^{[i]}(k)\|, \tag{26}$$

where we use $\sum_{j=1}^{N} a_j^i(k)=1$. Similarly, it holds that $\sum_{k=0}^{+\infty}\alpha(k)\|\hat{\lambda}(k)-v_{\lambda}^{[i]}(k)\|<+\infty$.

We now consider the evolution of $x^{[i]}(k)$. Recall that $v_x^{[i]}(k)\in X$. By Lemma 9.1 with $Z=X$, $z=v_x^{[i]}(k)-\alpha(k)\mathcal{S}_x^{[i]}(k)$ and $y=v_x^{[i]}(k)$, we have

$$\begin{aligned}\|x^{[i]}(k+1)-v_x^{[i]}(k)\|^2 &\leq \|v_x^{[i]}(k)-\alpha(k)\mathcal{S}_x^{[i]}(k)-v_x^{[i]}(k)\|^2\\ &-\|x^{[i]}(k+1)-(v_x^{[i]}(k)-\alpha(k)\mathcal{S}_x^{[i]}(k))\|^2,\end{aligned}$$

and thus $\|e_x^{[i]}(k)+\alpha(k)\mathcal{S}_x^{[i]}(k)\|\leq\alpha(k)\|\mathcal{S}_x^{[i]}(k)\|$. With this relation, from Lemma 9.2 with $Z=X$ and $d^{[i]}(k)=\mathcal{S}_x^{[i]}(k)$, the following holds for some $\gamma>0$ and $0<\beta<1$:

$$\|x^{[i]}(k)-\hat{x}(k)\| \leq N\gamma\beta^{k-1}\sum_{i=0}^{N}\|x^{[i]}(0)\| + 2N\gamma\sum_{\tau=0}^{k-1}\beta^{k-\tau}\alpha(\tau)\|\mathcal{S}_x^{[i]}(\tau)\|. \tag{27}$$

Multiplying both sides of (27) by $\alpha(k)M_\mu(k)$ and using (25), we obtain

$$\begin{aligned}\alpha(k)M_\mu(k)\|x^{[i]}(k)-\hat{x}(k)\| &\leq N\gamma\sum_{i=0}^{N}\|x^{[i]}(0)\|\alpha(k)M_\mu(k)\beta^{k-1}+2N\gamma\alpha(k)M_\mu(k)\\ &\times\sum_{\tau=0}^{k-1}\beta^{k-\tau}\alpha(\tau)(D_F+D_{G^+}M_\mu(\tau)+D_H M_\lambda(\tau)).\end{aligned}$$

Notice that the above inequalities hold for all $i\in V$. Then by employing the relation of $ab\leq\frac{1}{2}(a^2+b^2)$ and regrouping similar terms, we obtain

$$\begin{aligned}&\alpha(k)M_\mu(k)\max_{i\in V}\|x^{[i]}(k)-\hat{x}(k)\| \leq N\gamma\big(\frac{1}{2}\sum_{i=0}^{N}\|x^{[i]}(0)\|+(D_F+D_{G^+}+D_H)\sum_{\tau=0}^{k-1}\beta^{k-\tau}\big)\\ &\times\alpha(k)^2M_\mu^2(k)+\frac{1}{2}N\gamma\sum_{i=0}^{N}\|x^{[i]}(0)\|\beta^{2(k-1)}+N\gamma\sum_{\tau=0}^{k-1}\beta^{k-\tau}\alpha(\tau)^2(D_F+D_{G^+}M_\mu^2(\tau)+D_H M_\lambda^2(\tau)).\end{aligned}$$

Part (a) gives that $\{\alpha(k)^2M_\mu^2(k)\}$ is summable. Combining this fact with $\sum_{\tau=0}^{k-1}\beta^{k-\tau}\leq\sum_{k=0}^{+\infty}\beta^k=\frac{1}{1-\beta}$, then we can say that the first term on the right-hand side in the above estimate is summable. It is easy to check that the second term is also summable. It follows from Part (a) that $\lim_{k\to+\infty}\alpha(k)^2(D_F+D_{G^+}M_\mu^2(k)+D_HM_\lambda^2(k))=0$ and $\{\alpha(k)^2(D_F+D_{G^+}M_\mu^2(k)+D_HM_\lambda^2(k))\}$ is summable. Then Lemma 7 in [23] with $\gamma_\ell=N\gamma\alpha(\ell)^2(D_F+D_{G^+}M_\mu^2(\ell)+D_HM_\lambda^2(\ell))$ ensures

that the third term is summable. Therefore, the summability of $\{\alpha(k)M_\mu(k)\max_{i\in V}\|x^{[i]}(k)-\hat{x}(k)\|\}$ is guaranteed. Following the same lines in (26), one can show the summability of $\{\alpha(k)M_\mu(k)\|v_x^{[i]}(k)-\hat{x}(k)\|\}$. Following analogous arguments, we have that $\{\alpha(k)M_\lambda(k)\|v_x^{[i]}(k)-\hat{x}(k)\|\}$ and $\{\alpha(k)\|v_x^{[i]}(k)-\hat{x}(k)\|\}$ are summable. ■

Remark 5.1: In Lemma 5.6, the assumption of all local constraint sets being identical is utilized to find an upper bound of the convergence rate of $\|\hat{x}(k)-v_x^{[i]}(k)\|$ to zero. This property is crucial to establish the summability of expansions pertaining to $\|\hat{x}(k)-v_x^{[i]}(k)\|$ in part (b).•

The following is a basic iteration relation of the DPPDS algorithm.

Lemma 5.7 (Basic iteration relation): The following estimates hold for any $x\in X$ and $(\mu,\lambda)\in\mathbb{R}^m_{\geq 0}\times\mathbb{R}^\nu_{\geq 0}$:

$$\begin{aligned}\sum_{i=1}^N\|e_x^{[i]}(k)+\alpha(k)\mathcal{S}_x^{[i]}(k)\|^2 &\leq \sum_{i=1}^N\alpha(k)^2\|\mathcal{S}_x^{[i]}(k)\|^2\\ &-\sum_{i=1}^N 2\alpha(k)(\mathcal{H}^{[i]}(v_x^{[i]}(k),v_\mu^{[i]}(k),v_\lambda^{[i]}(k))-\mathcal{H}^{[i]}(x,v_\mu^{[i]}(k),v_\lambda^{[i]}(k)))\\ &+\sum_{i=1}^N(\|x^{[i]}(k)-x\|^2-\|x^{[i]}(k+1)-x\|^2),\end{aligned}\tag{28}$$

and,

$$\begin{aligned}0\leq\sum_{i=1}^N(\|\mu^{[i]}(k)-\mu\|^2-\|\mu^{[i]}(k+1)-\mu\|^2)+\sum_{i=1}^N(\|\lambda^{[i]}(k)-\lambda\|^2-\|\lambda^{[i]}(k+1)-\lambda\|^2)+\\ \sum_{i=1}^N 2\alpha(k)(\mathcal{H}^{[i]}(v_x^{[i]}(k),v_\mu^{[i]}(k),v_\lambda^{[i]}(k))-\mathcal{H}^{[i]}(v_x^{[i]}(k),\mu,\lambda))+\sum_{i=1}^N\alpha(k)^2(\|[g(v_x^{[i]}(k))]^+\|^2+\|h(v_x^{[i]}(k))\|^2).\end{aligned}\tag{29}$$

Proof: One can finish the proof by following analogous arguments in Lemma 5.3. ■

Lemma 5.8 (Achieving consensus): Let us suppose that the non-degeneracy assumption 2.2, the balanced communication assumption 2.3 and the periodical strong connectivity assumption 2.4 hold. Consider the sequences of $\{x^{[i]}(k)\}$, $\{\mu^{[i]}(k)\}$, $\{\lambda^{[i]}(k)\}$ and $\{y^{[i]}(k)\}$ of the distributed penalty primal-dual subgradient algorithm with the step-size sequence $\{\alpha(k)\}$ and the associated $\{s(k)\}$ satisfying $\lim_{k\rightarrow+\infty}\alpha(k)=0$ and $\lim_{k\rightarrow+\infty}\alpha(k+1)s(k)=0$. Then there exists $\tilde{x}\in X$ such that $\lim_{k\rightarrow+\infty}\|x^{[i]}(k)-\tilde{x}\|=0$ for all $i\in V$. Furthermore, $\lim_{k\rightarrow+\infty}\|\mu^{[i]}(k)-\mu^{[j]}(k)\|=0$, $\lim_{k\rightarrow+\infty}\|\lambda^{[i]}(k)-\lambda^{[j]}(k)\|=0$ and $\lim_{k\rightarrow+\infty}\|y^{[i]}(k)-y^{[j]}(k)\|=0$ for all $i,j\in V$.

Proof: Similar to (14), we have

$$\sum_{i=1}^{N}\|x^{[i]}(k+1)-x\|^2 \leq \sum_{i=1}^{N}\|x^{[i]}(k)-x\|^2+\sum_{i=1}^{N}\alpha(k)^2\|\mathcal{S}_x^{[i]}(k)\|^2+\sum_{i=1}^{N}2\alpha(k)\|\mathcal{S}_x^{[i]}(k)\|\|v_x^{[i]}(k)-x\|.$$

Since $\lim_{k\to+\infty}\alpha(k)\|\mathcal{S}_x^{[i]}(k)\|=0$, the proofs of $\lim_{k\to+\infty}\|x^{[i]}(k)-\tilde{x}\|=0$ for all $i\in V$ are analogous to those in Lemma 5.4. The remainder of the proofs can be finished by Proposition 9.1 with the properties of $\lim_{k\to+\infty}u_\mu^{[i]}(k)=0$, $\lim_{k\to+\infty}u_\lambda^{[i]}(k)=0$ and $\lim_{k\to+\infty}u_y^{[i]}(k)=0$ (due to $\lim_{k\to+\infty}x^{[i]}(k)=\tilde{x}$ and $f^{[i]}$ is continuous). ■

We now proceed to show Theorem 4.2 based on five claims.

Proof of Theorem 4.2:

Claim 1: For any $x^*\in X^*$ and $(\mu^*,\lambda^*)\in D_P^*$, the sequences of $\{\alpha(k)\big[\sum_{i=1}^{N}\mathcal{H}^{[i]}(x^*,v_\mu^{[i]}(k),v_\lambda^{[i]}(k))-\mathcal{H}(x^*,\hat{\mu}(k),\hat{\lambda}(k))\big]\}$ and $\{\alpha(k)\big[\sum_{i=1}^{N}\mathcal{H}^{[i]}(v_x^{[i]}(k),\mu^*,\lambda^*)-\mathcal{H}(\hat{x}(k),\mu^*,\lambda^*)\big]\}$ are summable.

Proof: Observe that

$$\begin{aligned}&\|\mathcal{H}^{[i]}(x^*,v_\mu^{[i]}(k),v_\lambda^{[i]}(k))-\mathcal{H}^{[i]}(x^*,\hat{\mu}(k),\hat{\lambda}(k))\|\\&\leq\|v_\mu^{[i]}(k)-\hat{\mu}(k)\|\|[g(x^*)]^+\|+\|v_\lambda^{[i]}(k)-\hat{\lambda}(k)\|\|h(x^*)\|\\&\leq G^+\|v_\mu^{[i]}(k)-\hat{\mu}(k)\|+H\|v_\lambda^{[i]}(k)-\hat{\lambda}(k)\|.\end{aligned}\qquad(30)$$

By using the summability of $\{\alpha(k)\|\hat{\mu}(k)-v_\mu^{[i]}(k)\|\}$ and $\{\alpha(k)\|\hat{\lambda}(k)-v_\lambda^{[i]}(k)\|\}$ in Part (b) of Lemma 5.6, we have that $\{\alpha(k)\sum_{i=1}^{N}\|\mathcal{H}^{[i]}(x^*,v_\mu^{[i]}(k),v_\lambda^{[i]}(k))-\mathcal{H}^{[i]}(x^*,\hat{\mu}(k),\hat{\lambda}(k))\|\}$ and thus $\{\alpha(k)\big[\sum_{i=1}^{N}\big(\mathcal{H}^{[i]}(x^*,v_\mu^{[i]}(k),v_\lambda^{[i]}(k))-\mathcal{H}^{[i]}(x^*,\hat{\mu}(k),\hat{\lambda}(k))\big)\big]\}$ are summable. Similarly, the following estimates hold:

$$\begin{aligned}&\|\mathcal{H}^{[i]}(v_x^{[i]}(k),\mu^*,\lambda^*)-\mathcal{H}^{[i]}(\hat{x}(k),\mu^*,\lambda^*)\|\leq\|f^{[i]}(v_x^{[i]}(k))-f^{[i]}(\hat{x}(k))\|\\&+\|(\mu^*)^T([g(v_x^{[i]}(k))]^+-[g(\hat{x}(k))]^+)\|+\|(\lambda^*)^T(|h(v_x^{[i]}(k))|-|h(\hat{x}(k))|)\|\\&\leq(D_F+D_{G^+}\|\mu^*\|+D_H\|\lambda^*\|)\|v_x^{[i]}(k)-\hat{x}(k)\|.\end{aligned}$$

Then the property of $\sum_{k=0}^{+\infty}\alpha(k)\|\hat{x}(k)-v_x^{[i]}(k)\|<+\infty$ in Part (b) of Lemma 5.6 implies the summability of the sequence $\{\alpha(k)\sum_{i=1}^{N}\|\mathcal{H}^{[i]}(v_x^{[i]}(k),\mu^*,\lambda^*)-\mathcal{H}^{[i]}(\hat{x}(k),\mu^*,\lambda^*)\|\}$ and thus the sequence $\{\alpha(k)\sum_{i=1}^{N}\big(\mathcal{H}^{[i]}(v_x^{[i]}(k),\mu^*,\lambda^*)-\mathcal{H}^{[i]}(\hat{x}(k),\mu^*,\lambda^*)\big)\}$. ■

Claim 2: Denote the weighted version of the local penalty function $\mathcal{H}^{[i]}$ over $[0,k-1]$ as

$$\hat{\mathcal{H}}^{[i]}(k):=\frac{1}{s(k-1)}\sum_{\ell=0}^{k-1}\alpha(\ell)\mathcal{H}^{[i]}(v_x^{[i]}(\ell),v_\mu^{[i]}(\ell),v_\lambda^{[i]}(\ell)).$$

The following property holds: $\lim_{k\to+\infty}\sum_{i=1}^{N}\hat{\mathcal{H}}^{[i]}(k)=p^*$.

Proof: Summing (28) over $[0,k-1]$ and replacing x by $x^*\in X^*$ leads to

$$\sum_{\ell=0}^{k-1}\alpha(\ell)\sum_{i=1}^{N}(\mathcal{H}^{[i]}(v_x^{[i]}(\ell),v_\mu^{[i]}(\ell),v_\lambda^{[i]}(\ell))-\mathcal{H}^{[i]}(x^*,v_\mu^{[i]}(\ell),v_\lambda^{[i]}(\ell)))$$
$$\leq\sum_{i=1}^{N}\|x^{[i]}(0)-x^*\|^2+\sum_{\ell=0}^{k-1}\sum_{i=1}^{N}\alpha(\ell)^2\|\mathcal{S}_x^{[i]}(\ell)\|^2. \quad (31)$$

The summability of $\{\alpha(k)^2\|\mathcal{S}_x^{[i]}(k)\|^2\}$ in Part (b) of Lemma 5.6 implies that the right-hand side of (31) is finite as $k\to+\infty$, and thus

$$\limsup_{k\to\infty}\frac{1}{s(k-1)}\sum_{\ell=0}^{k-1}\alpha(\ell)\big[\sum_{i=1}^{N}\big(\mathcal{H}^{[i]}(v_x^{[i]}(\ell),v_\mu^{[i]}(\ell),v_\lambda^{[i]}(\ell))-\mathcal{H}^{[i]}(x^*,v_\mu^{[i]}(\ell),v_\lambda^{[i]}(\ell))\big)\big]\leq 0. \quad (32)$$

Pick any $(\mu^*,\lambda^*)\in D_P^*$. It follows from Theorem 4.1 that (x^*,μ^*,λ^*) is a saddle point of $\mathcal{H}$ over $X\times\mathbb{R}^m_{\geq 0}\times\mathbb{R}^\nu_{\geq 0}$. Since $(\hat{\mu}(k),\hat{\lambda}(k))\in\mathbb{R}^m_{\geq 0}\times\mathbb{R}^\nu_{\geq 0}$, then we have $\mathcal{H}(x^*,\hat{\mu}(k),\hat{\lambda}(k))\leq\mathcal{H}(x^*,\mu^*,\lambda^*)=p^*$. Combining this relation, Claim 1 and (32) renders that

$$\limsup_{k\to+\infty}\frac{1}{s(k-1)}\sum_{\ell=0}^{k-1}\alpha(\ell)\big[\sum_{i=1}^{N}\mathcal{H}^{[i]}(v_x^{[i]}(\ell),v_\mu^{[i]}(\ell),v_\lambda^{[i]}(\ell))-p^*\big]$$
$$\leq\limsup_{k\to+\infty}\frac{1}{s(k-1)}\sum_{\ell=0}^{k-1}\alpha(\ell)\big[\sum_{i=1}^{N}\big(\mathcal{H}^{[i]}(v_x^{[i]}(\ell),v_\mu^{[i]}(\ell),v_\lambda^{[i]}(\ell))-\mathcal{H}^{[i]}(x^*,v_\mu^{[i]}(\ell),v_\lambda^{[i]}(\ell))\big)\big]$$
$$+\limsup_{k\to+\infty}\frac{1}{s(k-1)}\sum_{\ell=0}^{k-1}\alpha(\ell)\big[\sum_{i=1}^{N}\mathcal{H}^{[i]}(x^*,v_\mu^{[i]}(\ell),v_\lambda^{[i]}(\ell))-\mathcal{H}(x^*,\hat{\mu}(\ell),\hat{\lambda}(\ell))\big]$$
$$+\limsup_{k\to+\infty}\frac{1}{s(k-1)}\sum_{\ell=0}^{k-1}(\mathcal{H}(x^*,\hat{\mu}(\ell),\hat{\lambda}(\ell))-p^*)\leq 0,$$

and thus $\limsup_{k\to+\infty}\sum_{i=1}^{N}\hat{\mathcal{H}}^{[i]}(k)\leq p^*$.

On the other hand, $\hat{x}(k)\in X$ (due to the fact that X is convex) implies that $\mathcal{H}(\hat{x}(k),\mu^*,\lambda^*)\geq\mathcal{H}(x^*,\mu^*,\lambda^*)=p^*$. Along similar lines, by using (29) with $\mu=\mu^*$, $\lambda=\lambda^*$, and Claim 1, we have the following estimate: $\liminf_{k\to+\infty}\sum_{i=1}^{N}\hat{\mathcal{H}}^{[i]}(k)\geq p^*$. Then we have the desired relation. ■

Claim 3: Denote by $\pi(k):=\sum_{i=1}^{N}\mathcal{H}^{[i]}(v_x^{[i]}(k),v_\mu^{[i]}(k),v_\lambda^{[i]}(k))-\mathcal{H}(\hat{x}(k),\hat{\mu}(k),\hat{\lambda}(k))$. And we denote the weighted version of the global penalty function $\mathcal{H}$ over $[0,k-1]$ as

$$\hat{\mathcal{H}}(k):=\frac{1}{s(k-1)}\sum_{\ell=0}^{k-1}\alpha(\ell)\mathcal{H}(\hat{x}(\ell),\hat{\mu}(\ell),\hat{\lambda}(\ell)).$$

The following property holds: $\lim\limits_{k\to+\infty}\hat{\mathcal{H}}(k)=p^*$.

Proof: Notice that

$$\pi(k)=\sum_{i=1}^{N}(f^{[i]}(v_x^{[i]}(k))-f^{[i]}(\hat{x}(k)))+\sum_{i=1}^{N}\big(v_\mu^{[i]}(k)^T[g(v_x^{[i]}(k))]^+-v_\mu^{[i]}(k)^T[g(\hat{x}(k))]^+\big)$$
$$+\sum_{i=1}^{N}\big(v_\mu^{[i]}(k)^T[g(\hat{x}(k))]^+-\hat{\mu}(k)^T[g(\hat{x}(k))]^+\big)+\sum_{i=1}^{N}\big(v_\lambda^{[i]}(k)^T|h(v_x^{[i]}(k))|-v_\lambda^{[i]}(k)^T|h(\hat{x}(k))|\big)$$
$$+\sum_{i=1}^{N}\big(v_\lambda^{[i]}(k)^T|h(\hat{x}(k))|-\hat{\lambda}(k)^T|h(\hat{x}(k))|\big). \tag{33}$$

By using the boundedness of subdifferentials and the primal estimates, it follows from (33) that

$$\|\pi(k)\|\leq(D_F+D_{G^+}M_\mu(k)+D_HM_\lambda(k))\times\sum_{i=1}^{N}\|v_x^{[i]}(k)-\hat{x}(k)\|$$
$$+G^+\sum_{i=1}^{N}\|v_\mu^{[i]}(k)-\hat{\mu}(k)\|+H\sum_{i=1}^{N}\|v_\lambda^{[i]}(k)-\hat{\lambda}(k)\|. \tag{34}$$

Then it follows from (b) in Lemma 5.6 that $\{\alpha(k)\|\pi(k)\|\}$ is summable. Notice that $\|\hat{\mathcal{H}}(k)-\sum_{i=1}^{N}\hat{\mathcal{H}}^{[i]}(k)\|\leq\frac{\sum_{\ell=0}^{k-1}\alpha(\ell)\|\pi(\ell)\|}{s(k-1)}$, and thus $\lim\limits_{k\to+\infty}\|\hat{\mathcal{H}}(k)-\sum\limits_{i=1}^{N}\hat{\mathcal{H}}^{[i]}(k)\|=0$. The desired result immediately follows from Claim 2. ∎

Claim 4: The limit point $\tilde{x}$ in Lemma 5.8 is a primal optimal solution.

Proof: Let $\hat{\mu}(k)=(\hat{\mu}_1(k),\cdots,\hat{\mu}_m(k))^T\in\mathbb{R}^m_{\geq0}$. By the balanced communication assumption 2.3, we obtain

$$\sum_{i=1}^{N}\mu^{[i]}(k+1)=\sum_{i=1}^{N}\sum_{j=1}^{N}a_j^i(k)\mu^{[j]}(k)+\alpha(k)\sum_{i=1}^{N}[g(v_x^{[i]}(k))]^+$$
$$=\sum_{j=1}^{N}\mu^{[j]}(k)+\alpha(k)\sum_{i=1}^{N}[g(v_x^{[i]}(k))]^+.$$

This implies that the sequence $\{\hat{\mu}_\ell(k)\}$ is non-decreasing in $\mathbb{R}_{\geq0}$. Observe that $\{\hat{\mu}_\ell(k)\}$ is lower bounded by zero. In this way, we distinguish the following two cases:

Case 1: The sequence $\{\hat{\mu}_\ell(k)\}$ is upper bounded. Then $\{\hat{\mu}_\ell(k)\}$ is convergent in $\mathbb{R}_{\geq0}$. Recall that $\lim\limits_{k\to+\infty}\|\mu^{[i]}(k)-\mu^{[j]}(k)\|=0$ for all $i,j\in V$. This implies that there exists $\mu_\ell^*\in\mathbb{R}_{\geq0}$ such that $\lim\limits_{k\to+\infty}\|\mu_\ell^{[i]}(k)-\mu_\ell^*\|=0$ for all $i\in V$. Observe that $\sum_{i=1}^{N}\mu^{[i]}(k+1)=\sum_{i=1}^{N}\mu^{[i]}(0)+\sum_{\tau=0}^{k}\alpha(\tau)\sum_{i=1}^{N}[g(v_x^{[i]}(\tau))]^+$. Thus, we have $\sum_{k=0}^{+\infty}\alpha(k)\sum_{i=1}^{N}[g_\ell(v_x^{[i]}(k))]^+<+\infty$, implying

that $\liminf_{k\to+\infty}[g_\ell(v_x^{[i]}(k))]^+ = 0$. Since $\lim_{k\to+\infty}\|x^{[i]}(k)-\tilde{x}\| = 0$ for all $i\in V$, then $\lim_{k\to+\infty}\|v_x^{[i]}(k)-\tilde{x}\| = 0$, and thus $[g_\ell(\tilde{x})]^+ = 0$.

<u>Case 2:</u> The sequence $\{\hat{\mu}_\ell(k)\}$ is not upper bounded. Since $\{\hat{\mu}_\ell(k)\}$ is non-decreasing, then $\hat{\mu}_\ell(k) \to +\infty$. It follows from Claim 3 and (a) in Lemma 5.1 that it is impossible that $\mathcal{H}(\hat{x}(k),\hat{\mu}(k),\hat{\lambda}(k)) \to +\infty$. Assume that $[g_\ell(\tilde{x})]^+ > 0$. Then we have

$$\begin{aligned}\mathcal{H}(\hat{x}(k),\hat{\mu}(k),\hat{\lambda}(k)) &= f(\hat{x}(k)) + N\hat{\mu}(k)^T[g(\hat{x}(k))]^+ + N\lambda(k)^T|h(\hat{x}(k))|\\ &\geq f(\hat{x}(k)) + \hat{\mu}_\ell(k)[g_\ell(\hat{x}(k))]^+.\end{aligned} \tag{35}$$

Taking limits on both sides of (35) and we obtain:

$$\liminf_{k\to+\infty}\mathcal{H}(\hat{x}(k),\hat{\mu}(k),\hat{\lambda}(k)) \geq \limsup_{k\to+\infty}(f(\hat{x}(k)) + \hat{\mu}_\ell(k)[g_\ell(\hat{x}(k))]^+) = +\infty.$$

Then we reach a contradiction, implying that $[g_\ell(\tilde{x})]^+ = 0$.

In both cases, we have $[g_\ell(\tilde{x})]^+ = 0$ for any $1\leq \ell\leq m$. By utilizing similar arguments, we can further prove that $|h(\tilde{x})| = 0$. Since $\tilde{x}\in X$, then $\tilde{x}$ is feasible and thus $f(\tilde{x})\geq p^*$. On the other hand, since $\frac{\sum_{\ell=0}^{k-1}\alpha(\ell)\hat{x}(\ell)}{\sum_{\ell=0}^{k-1}\alpha(\ell)}$ is a convex combination of $\hat{x}(0),\cdots,\hat{x}(k-1)$ and $\lim_{k\to+\infty}\hat{x}(k) = \tilde{x}$, then Claim 3 and (b) in Lemma 5.1 implies that

$$p^* = \lim_{k\to+\infty}\hat{\mathcal{H}}(k) = \lim_{k\to+\infty}\frac{\sum_{\ell=0}^{k-1}\alpha(\ell)\mathcal{H}(\hat{x}(\ell),\hat{\mu}(\ell),\hat{\lambda}(\ell))}{\sum_{\ell=0}^{k-1}\alpha(\ell)} \geq \lim_{k\to+\infty} f(\frac{\sum_{\ell=0}^{k-1}\alpha(\ell)\hat{x}(\ell)}{\sum_{\ell=0}^{k-1}\alpha(\ell)}) = f(\tilde{x}).$$

Hence, we have $f(\tilde{x}) = p^*$ and thus $\tilde{x}\in X^*$. ■

Claim 5: It holds that $\lim_{k\to+\infty}\|y^{[i]}(k)-p^*\| = 0$.

Proof: The proof follows the same lines in Claim 2 of Theorem 3.2 and thus omitted here.

■

VI. Discussion

In this section, we present some possible extensions and interesting special cases.

A. Discussion on the periodic strong connectivity assumption in Theorem 3.2

In the case that $\mathcal{G}(k)$ is undirected, then the periodic strong connectivity assumption 2.4 in Theorem 3.2 can be weakened into:

Assumption 6.1 (Eventual strong connectivity): The undirected graph $(V,\cup_{k\geq s}E(k))$ is connected for all time instant $s\geq 0$.

If $\mathcal{G}(k)$ is undirected, the periodic connectivity assumption 2.4 in Theorem 3.2 can also be replaced with the assumption in Proposition 2 of [18]; i.e., for any time instant $k \geq 0$, there is an agent connected to all other agents in the undirected graph $(V, \cup_{k\geq s} E(k))$.

B. A generalized step-size scheme

The step-size scheme in the DLPDS algorithm can be slightly generalized the case that the maximum deviation of step-sizes between agents at each time is not large. It is formally stated as follows: $\lim_{k\to+\infty} \alpha^{[i]}(k) = 0$, $\sum_{k=0}^{+\infty} \alpha^{[i]}(k) = +\infty$, $\sum_{k=0}^{+\infty} \alpha^{[i]}(k)^2 < +\infty$, $\min_{i\in V} \alpha^{[i]}(k) \geq C_\alpha \max_{i\in V} \alpha^{[i]}(k)$, where $\alpha^{[i]}(k)$ is the step-size of agent i at time k and $C_\alpha \in (0, 1]$.

C. Discussion on the Slater's condition in Theorem 4.2

If g_ℓ ($1 \leq \ell \leq m$) is linear, then the Slater's condition 2.1 can be weakened to the following: there exists a relative interior point $\bar{x}$ of X such that $h(\bar{x}) = 0$ and $g(\bar{x}) \leq 0$. For this case, the strong duality and the non-emptyness of the penalty dual optimal set can be ensured by replacing Proposition 5.3.5 [3] with Proposition 5.3.4 [3] in the proofs of Lemma 4.1. In this way, the convergence results of the DPPDS algorithm still hold for the case of linear g_ℓ.

D. The special case in the absence of inequality and equality constraints

The following special case of problem (1) is studied in [23]:

$$\min_{x\in\mathbb{R}^n} \sum_{i=1}^{N} f^{[i]}(x), \quad \text{s.t.} \quad x \in \cap_{i=1}^{N} X^{[i]}. \tag{36}$$

In order to solve problem (36), we consider the following *Distributed Primal Subgradient* Algorithm which is a special case of the DLPDS algorithm:

$$x^{[i]}(k+1) = P_{X^{[i]}}[v_x^{[i]}(k) - \alpha(k)\mathcal{D}f^{[i]}(v_x^{[i]}(k))].$$

Corollary 6.1 (Convergence properties of the distributed primal subgradient algorithm): Consider problem (36), and let the non-degeneracy assumption 2.2, the balanced communication assumption 2.3 and the periodic strong connectivity assumption 2.4 hold. Consider the sequence $\{x^{[i]}(k)\}$ of the distributed primal subgradient algorithm with initial states $x^{[i]}(0) \in X^{[i]}$ and the step-sizes satisfying $\lim_{k\to+\infty} \alpha(k) = 0$, $\sum_{k=0}^{+\infty} \alpha(k) = +\infty$, and $\sum_{k=0}^{+\infty} \alpha(k)^2 < +\infty$. Then there exists an optimal solution x^* such that $\lim_{k\to+\infty} \|x^{[i]}(k) - x^*\| = 0$ all $i \in V$.

Proof: The result is an immediate consequence of Theorem 3.2 with $g(x) \equiv 0$. ■

VII. NUMERICAL EXAMPLES

In this section, we illustrate the performance of the DLPDS and DPPDS algorithms via two numerical examples.

A. A numerical example of NUM for the DLPDS algorithm

In order to study the performance of the DLPDS algorithm, we here consider a numerical example of network utility maximization, e.g., in [15]. Consider five agents and one link where each agent sends data through the link at a rate of z_i, and the link capacity is 5. The global decision vector $x := [z_1 \cdots z_5]^T$ is the resource allocation vector. Each agent i is associated a concave utility function $f^{[i]}(z_i) := \sqrt{z_i}$, representing the utility agent i obtains through sending data at a rate of z_i. Agents aim to maximize the aggregate sum of local utilities and this problem can be formulated as follows:

$$\min_{x \in \mathbb{R}^5} \sum_{i \in V} -\sqrt{z_i} \quad \text{s.t.} \quad z_1 + z_2 + z_3 + z_4 + z_5 \leq 5, \quad x \in \cap_{i \in V} X^{[i]}, \tag{37}$$

where local constraint sets $X^{[i]}$ are given by:

$$X^{[1]} := [0.5,\ \ 5.5] \times [0.5,\ \ 5.5] \times [0.5,\ \ 5.5] \times [0.5,\ \ 5.5] \times [0.5,\ \ 5.5],$$

$$X^{[2]} := [0.55,\ \ 5.25] \times [0.55,\ \ 5.25] \times [0.55,\ \ 5.25] \times [0.55,\ \ 5.25] \times [0.55,\ \ 5.25],$$

$$X^{[3]} := [0.5,\ \ 6] \times [0.5,\ \ 6] \times [0.5,\ \ 6] \times [0.5,\ \ 6] \times [0.5,\ \ 6],$$

$$X^{[4]} := [0.5,\ \ 5] \times [0.5,\ \ 5] \times [0.5,\ \ 5] \times [0.5,\ \ 5] \times [0.5,\ \ 5],$$

$$X^{[5]} := [0.525,\ \ 5.75] \times [0.525,\ \ 5.75] \times [0.525,\ \ 5.75] \times [0.525,\ \ 5.75] \times [0.525,\ \ 5.75].$$

We use the DLPDS algorithm to solve problem (37) by choosing step-size $\alpha(k) = \frac{1}{k+1}$. Figures 1 to 5 shows the simulation results of the DLPDS algorithm in comparison with the centralized subgradient algorithm. It demonstrates that all the agents takes 10^4 iterates to agree upon the optimal solution $[1\ \ 1\ \ 1\ \ 1\ \ 1]^T$. Furthermore, it can be observed that the optimal solution can be found by the centralized subgradient algorithm with the same step-size after 200 iterates which is much less than that of the DLPDS algorithm.

B. A numerical example for the DPPDS algorithm

Consider a network with five agents and their objective functions are defined as

$$
\begin{aligned}
f^{[1]}(x) &:= \frac{1}{5}\big((a-5)^2 + (b-2.5)^2 + (c-5)^2 + (d+2.5)^2 + (e+5)^2\big),\\
f^{[2]}(x) &:= \frac{1}{5}\big((a-2.5)^2 + (b-5)^2 + (c+2.5)^2 + (d+5)^2 + (e-5)^2\big),\\
f^{[3]}(x) &:= \frac{1}{5}\big((a-5)^2 + (b+2.5)^2 + (c+5)^2 + (d-5)^2 + (e-2.5)^2\big),\\
f^{[4]}(x) &:= \frac{1}{5}\big((a+2.5)^2 + (b+5)^2 + (c-5)^2 + (d-2.5)^2 + (e-5)^2\big),\\
f^{[5]}(x) &:= \frac{1}{5}\big((a+5)^2 + (b-5)^2 + (c-2.5)^2 + (d-5)^2 + (e+2.5)^2\big),
\end{aligned}
$$

where the global decision vector $x := [a \quad b \quad c \quad d \quad e]^T \in \mathbb{R}^5$. The global equality constraint function is given by $h(x) := a+b+c+d+e-5$, and the global constraint set is as follows: $X := [-5 \quad 5] \times [-5 \quad 5] \times [-5 \quad 5] \times [-5 \quad 5] \times [-5 \quad 5]$. Consider the optimization problem as follows:

$$
\min_{x\in\mathbb{R}^5} \sum_{i\in V} f^{[i]}(x), \quad \text{s.t.} \quad h(x) = 0, \quad x \in X.
$$

We employ the DPPDS algorithm to solve the above optimization problem with the step-size $\alpha(k) = \frac{1}{k+1}$. Its simulation results are included in Figures 6 to 10 in comparison with the performance of the centralized subgradient algorithm. Observe that all the agents asymptotically achieve the optimal solution $[1 \quad 1 \quad 1 \quad 1 \quad 1]^T$. Like the DLPDS algorithm, convergence rate of the DPPDS algorithm is slower than the centralized algorithm.

VIII. Conclusion

We have studied a multi-agent optimization problem where the agents aim to minimize a sum of local objective functions subject to a global inequality constraint, a global equality constraint and a global constraint set defined as the intersection of local constraint sets. We have considered two cases: the first one in the absence of the equality constraint and the second one with identical local constraint sets. To address these cases, we have introduced two distributed subgradient algorithms which are based on Lagrangian and penalty primal-dual methods, respectively. These two algorithms were shown to asymptotically converge to primal solutions and optimal values. Two numerical examples were presented to demonstrate the performance our algorithms. Our future work includes explicit characterization of convergence rates of the algorithms in this paper.

IX. APPENDIX

A. Dynamic average consensus algorithms

The following is the vector version of the first-order dynamic average consensus algorithm proposed in [35] with $x^{[i]}(k), \xi^{[i]}(k) \in \mathbb{R}^n$:

$$x^{[i]}(k+1) = \sum_{j=1}^{N} a_j^i(k) x^{[j]}(k) + \xi^{[i]}(k). \tag{38}$$

Proposition 9.1: Denote $\Delta\xi_\ell(k) := \max_{i\in V} \xi_\ell^{[i]}(k) - \min_{i\in V} \xi_\ell^{[i]}(k)$ for $1 \leq \ell \leq n$. Let the non-degeneracy assumption 2.2, the balanced communication assumption 2.3 and the periodic strong connectivity assumption 2.4 hold. Assume that $\lim\limits_{k\to+\infty} \Delta\xi_\ell(k) = 0$ for all $1 \leq \ell \leq n$ and all $k \geq 0$. Then $\lim\limits_{k\to+\infty} \|x^{[i]}(k) - x^{[j]}(k)\| = 0$ for all $i, j \in V$.

B. A property of projection operators

The proof of the following lemma can be found in [3], [4] and [23].

Lemma 9.1: Let Z be a non-empty, closed and convex set in $\mathbb{R}^n$. For any $z \in \mathbb{R}^n$, the following holds for any $y \in Z$: $\|P_Z[z] - y\|^2 \leq \|z - y\|^2 - \|P_Z[z] - z\|^2$.

C. Some properties of the distributed projected subgradient algorithm in [23]

Consider the following distributed projected subgradient algorithm proposed in [23]: $x^{[i]}(k+1) = P_Z[v_x^{[i]}(k) - \alpha(k)d^{[i]}(k)]$. Denote by $e^{[i]}(k) := P_Z[v_x^{[i]}(k) - \alpha(k)d^{[i]}(k)] - v_x^{[i]}(k)$. The following is a slight modification of Lemma 8 and its proof in [23].

Lemma 9.2: Let the non-degeneracy assumption 2.2, the balanced communication assumption 2.3 and the periodic strong connectivity assumption 2.4 hold. Suppose $Z \in \mathbb{R}^n$ is a closed and convex set. Then there exist $\gamma > 0$ and $\beta \in (0,1)$ such that

$$\|x^{[i]}(k) - \hat{x}(k)\| \leq N\gamma \sum_{\tau=0}^{k-1} \beta^{k-\tau}\{\alpha(\tau)\|d^{[i]}(\tau)\| + \|e^{[i]}(\tau) + \alpha(\tau)d^{[i]}(\tau)\|\} + N\gamma\beta^{k-1}\sum_{i=0}^{N}\|x^{[i]}(0)\|.$$

Suppose $\{d^{[i]}(k)\}$ is uniformly bounded for each $i \in V$, and $\sum_{k=0}^{+\infty}\alpha(k)^2 < +\infty$, then we have $\sum_{k=0}^{+\infty}\alpha(k)\max_{i\in V}\|x^{[i]}(k) - \hat{x}(k)\| < +\infty$.

REFERENCES

[1] K.J. Arrow, L. Hurwicz, and H. Uzawa. *Studies in linear and nonlinear programming*. Stanford University Press, 1958.

[2] D. P. Bertsekas and J. N. Tsitsiklis. *Parallel and Distributed Computation: Numerical Methods*. Athena Scientific, 1997.

[3] D.P. Bertsekas. *Convex optimization theory*. Anthena Scietific, 2009.

[4] D.P. Bertsekas, A. Nedic, and A. Ozdaglar. *Convex analysis and optimization*. Anthena Scietific, 2003.

[5] V. D. Blondel, J. M. Hendrickx, A. Olshevsky, and J. N. Tsitsiklis. Convergence in multiagent coordination, consensus, and flocking. In *IEEE Conf. on Decision and Control and European Control Conference*, pages 2996–3000, Seville, Spain, December 2005.

[6] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah. Randomized gossip algorithms. *IEEE Transactions on Information Theory*, 52(6):2508–2530, 2006.

[7] J. Cortés. Analysis and design of distributed algorithms for χ-consensus. In *IEEE Conf. on Decision and Control*, pages 3363–3368, San Diego, USA, December 2006.

[8] M. C. DeGennaro and A. Jadbabaie. Decentralized control of connectivity for multi-agent systems. In *IEEE Conf. on Decision and Control*, pages 3947–3952, San Diego, USA, Dec 2006.

[9] J. Derenick and J. Spletzer. Convex optimization strategies for coordinating large-scale robot formations. *IEEE Transactions on Robotics*, 23(6):1252–1259, 2007.

[10] J. Derenick, J. Spletzer, and M. Ani Hsieh. An optimal approach to collaborative target tracking with performance guarantees. *Journal of Intelligent and Robotic Systems*, 56(1-2):47–67, 2009.

[11] J. A. Fax and R. M. Murray. Information flow and cooperative control of vehicle formations. *IEEE Transactions on Automatic Control*, 49(9):1465–1476, 2004.

[12] A. Jadbabaie, J. Lin, and A. S. Morse. Coordination of groups of mobile autonomous agents using nearest neighbor rules. *IEEE Transactions on Automatic Control*, 48(6):988–1001, 2003.

[13] B. Johansson, T. Keviczky, M. Johansson, and K. H. Johansson. Subgradient methods and consensus algorithms for solving convex optimization problems. In *IEEE Conf. on Decision and Control*, pages 4185–4190, Cancun, Mexico, December 2008.

[14] A. Kashyap, T. Başar, and R. Srikant. Quantized consensus. *Automatica*, 43(7):1192–1203, 2007.

[15] F. P. Kelly, A. Maulloo, and D. Tan. Rate control in communication networks: Shadow prices, proportional fairness and stability. *Journal of the Operational Research Society*, 49(3):237–252, 1998.

[16] P. Martin and M. Egerstedt. Optimization of multi-agent motion programs with applications to robotic marionettes. In *Hybrid Systems: Computation and Control*, April 2009.

[17] M. Mehyar, D. Spanos, J. Pongsajapan, S. H. Low, and R. M. Murray. Asynchronous distributed averaging on communication networks. *IEEE/ACM Transactions on Networking*, 15(3):512–520, 2007.

[18] L. Moreau. Stability of multiagent systems with time-dependent communication links. *IEEE Transactions on Automatic Control*, 50(2):169–182, 2005.

[19] A. I. Mourikis and S. I. Roumeliotis. Optimal sensing strategies for mobile robot formations: Resource-constrained localization. In *Proceedings of Robotics: Science and Systems*, pages 281–288, Cambridge, USA, 2005.

[20] A. Nedic and A. Ozdaglar. Approximate primal solutions and rate analysis for dual subgradient methods. *SIAM Journal on Optimization*, 19(4):1757–1780, 2009.

[21] A. Nedic and A. Ozdaglar. Distributed subgradient methods for multi-agent optimization. *IEEE Transactions on Automatic Control*, 54(1):48–61, 2009.

[22] A. Nedic and A. Ozdaglar. Subgradient methods for saddle-point problems. *Journal of Optimization Theory and Applications*, 142(1):205–228, 2009.

[23] A. Nedic, A. Ozdaglar, and P.A. Parrilo. Constrained consensus and optimization in multi-agent networks. *IEEE Transactions on Automatic Control*, 55(4):922–938, 2010.

[24] R. D. Nowak. Distributed EM algorithms for density estimation and clustering in sensor networks. *IEEE Transactions on Signal Processing*, 51:2245–2253, 2003.

[25] R. Olfati-Saber and R. M. Murray. Consensus problems in networks of agents with switching topology and time-delays. *IEEE Transactions on Automatic Control*, 49(9):1520–1533, 2004.

[26] A. Oliveira, S. Soares, and L. Nepomuceno. Optimal active power dispatch combining network flow and interior point approaches. *IEEE Transactions on Power Systems*, 18(4):1235–1240, 2003.

[27] A. Olshevsky and J. N. Tsitsiklis. Convergence speed in distributed consensus and averaging. *SIAM Journal on Control and Optimization*, 48(1):33–55, 2009.

[28] M. G. Rabbat and R. D. Nowak. Decentralized source localization and tracking. In *IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, pages 921–924, May 2004.

[29] A. Rantzer. Using game theory for distributed control engineering. In *Games 2008, 3rd World Congress of the Game Theory Society*, 2008.

[30] S. Sundhar Ram, A. Nedic, and V. V. Veeravalli. Distributed and recursive parameter estimation in parametrized linear state-space models. *IEEE Transactions on Automatic Control*, 55(2):488–492, 2010.

[31] A. Tahbaz-Salehi and A. Jadbabaie. Consensus over random networks. *IEEE Transactions on Automatic Control*, 53(3):791–795, 2008.

[32] J. N. Tsitsiklis. *Problems in Decentralized Decision Making and Computation*. PhD thesis, Massachusetts Institute of Technology, November 1984. Available at `http://web.mit.edu/jnt/www/Papers/PhD-84-jnt.pdf`.

[33] H. Wei, H. Sasaki, J. Kubokawa, and R. Yokoyama. An interior point nonlinear programming for optimal power flow problems with a novel data structure. *IEEE Transactions on Power Systems*, 13:870–877, 1998.

[34] L. Xiao and S. Boyd. Fast linear iterations for distributed averaging. *Systems & Control Letters*, 53:65–78, 2004.

[35] M. Zhu and S. Martínez. Discrete-time dynamic average consensus. *Automatica*, 46(2):322–329, 2010.

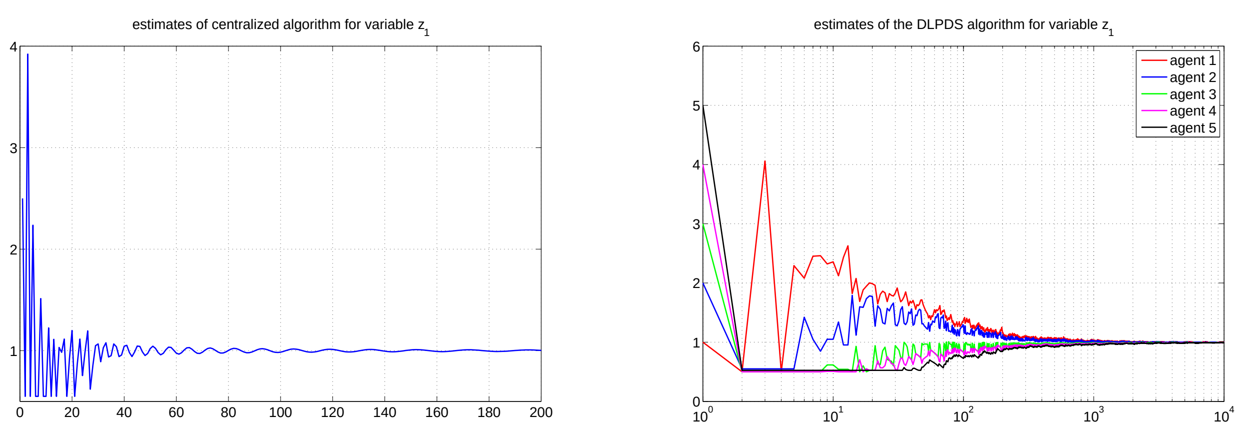


Fig. 1. Estimates of variable z_1 of centralized algorithm and the DLPDS algorithm

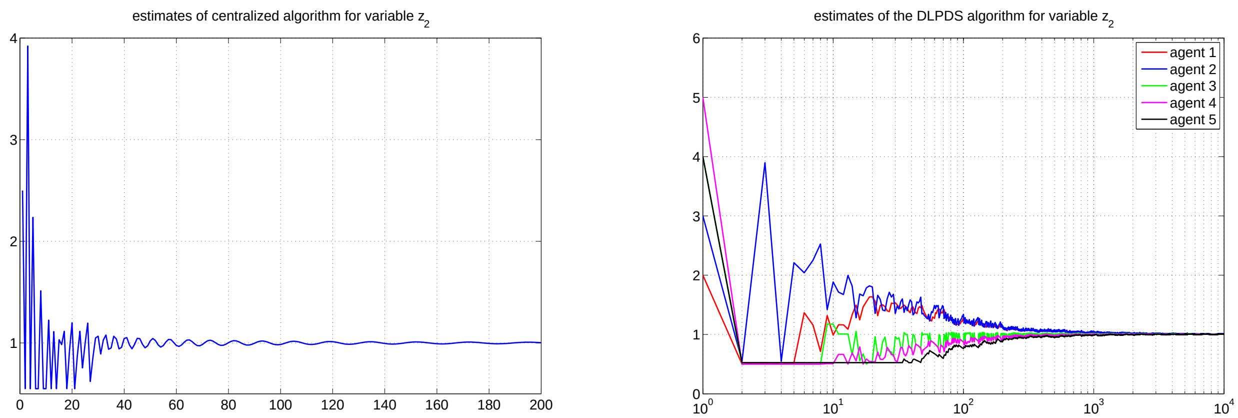


Fig. 2. Estimates of variable z_2 of centralized algorithm and the DLPDS algorithm

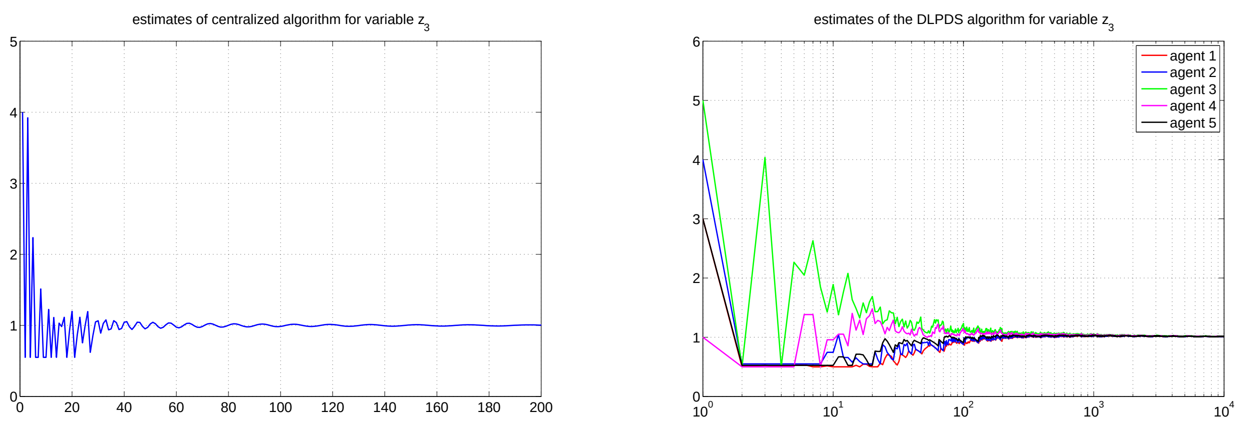


Fig. 3. Estimates of variable z_3 of centralized algorithm and the DLPDS algorithm

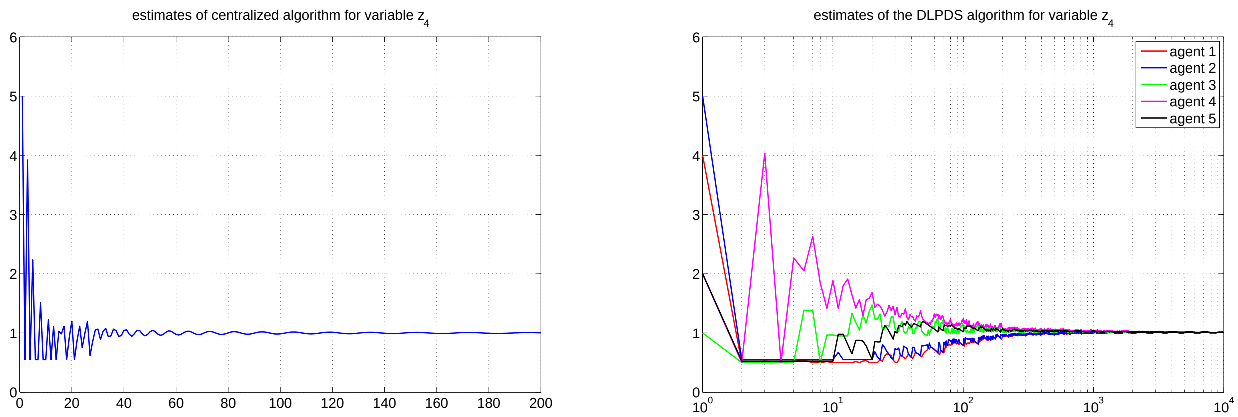


Fig. 4. Estimates of variable z_4 of centralized algorithm and the DLPDS algorithm

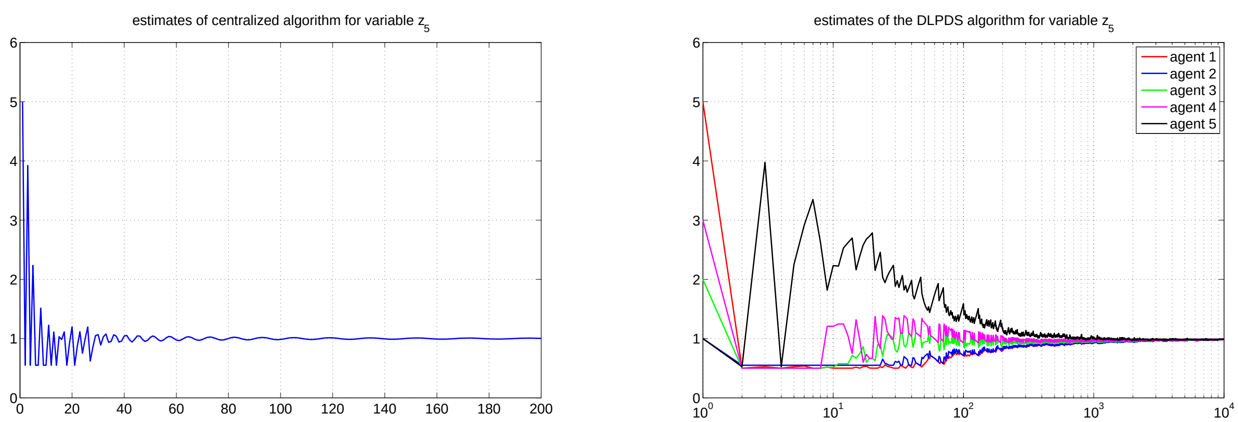


Fig. 5. Estimates of variable z_5 of centralized algorithm and the DLPDS algorithm

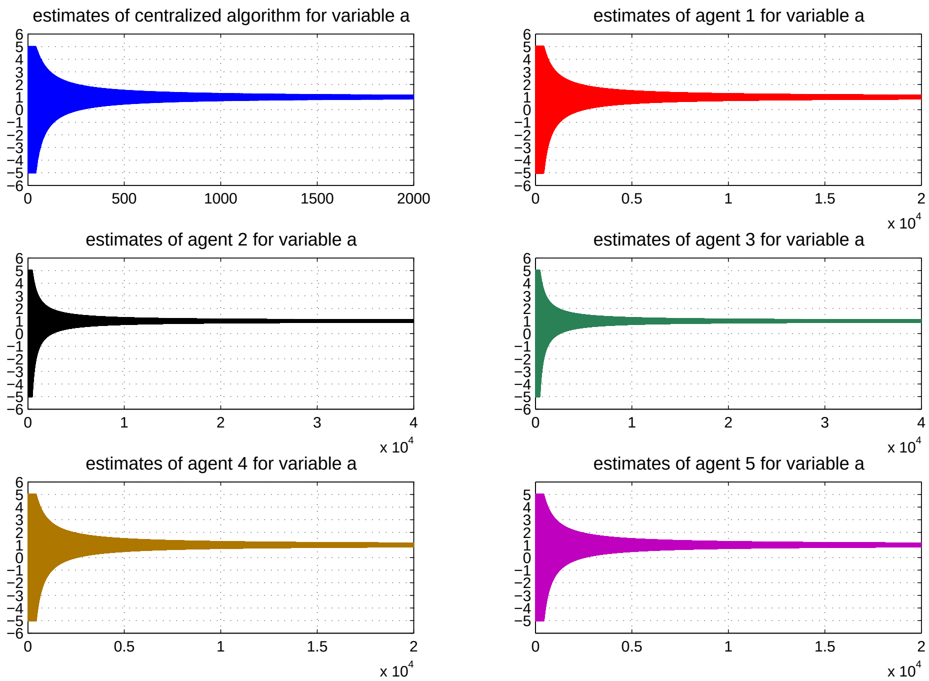


Fig. 6. Estimates of variable a in the DPPDS algorithm

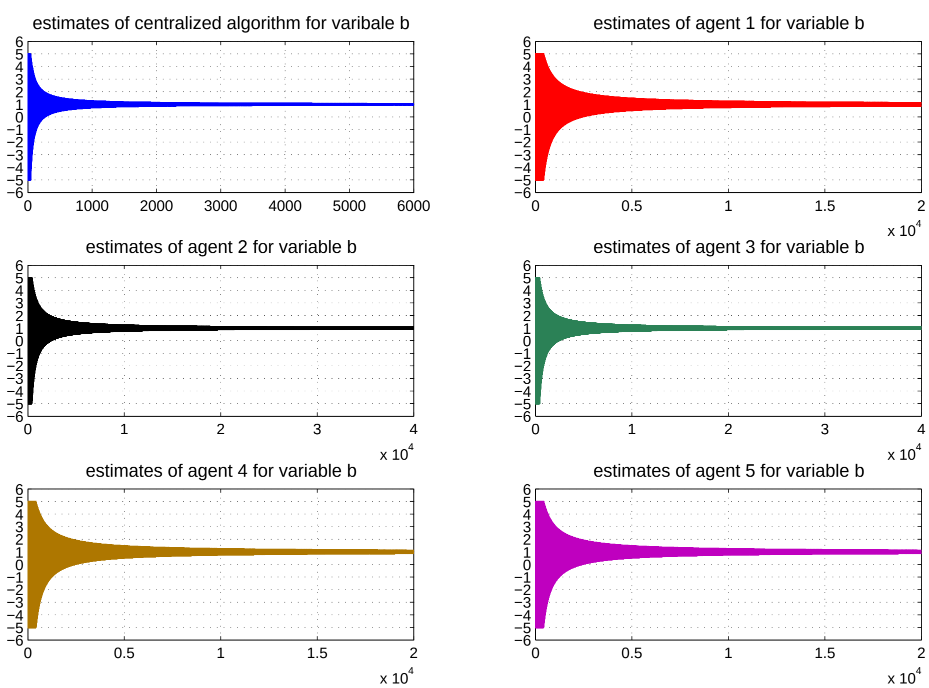


Fig. 7. Estimates of variable b in the DPPDS algorithm

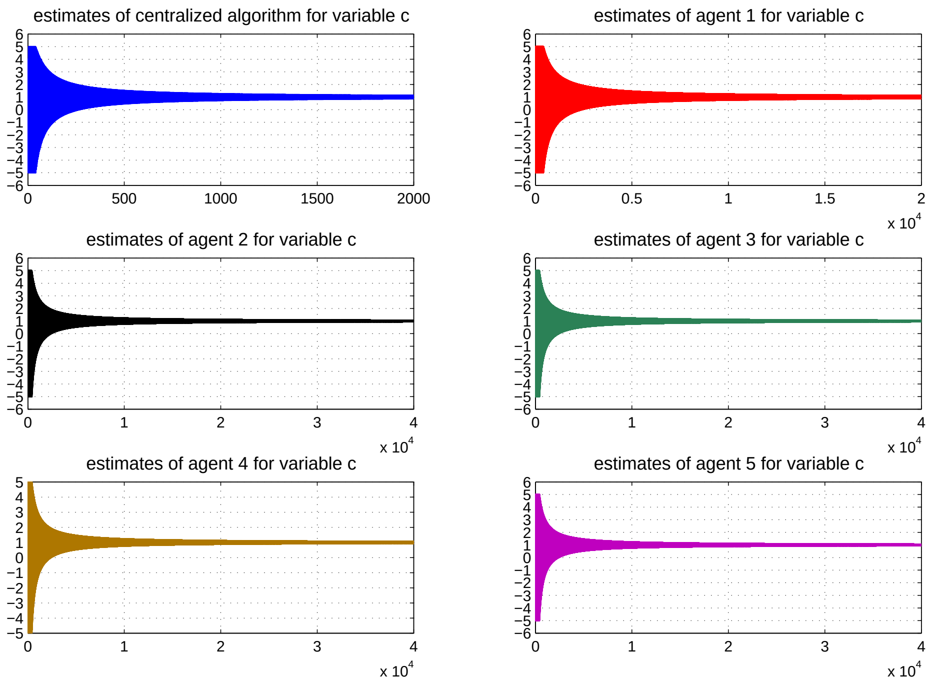


Fig. 8. Estimates of variable c in the DPPDS algorithm

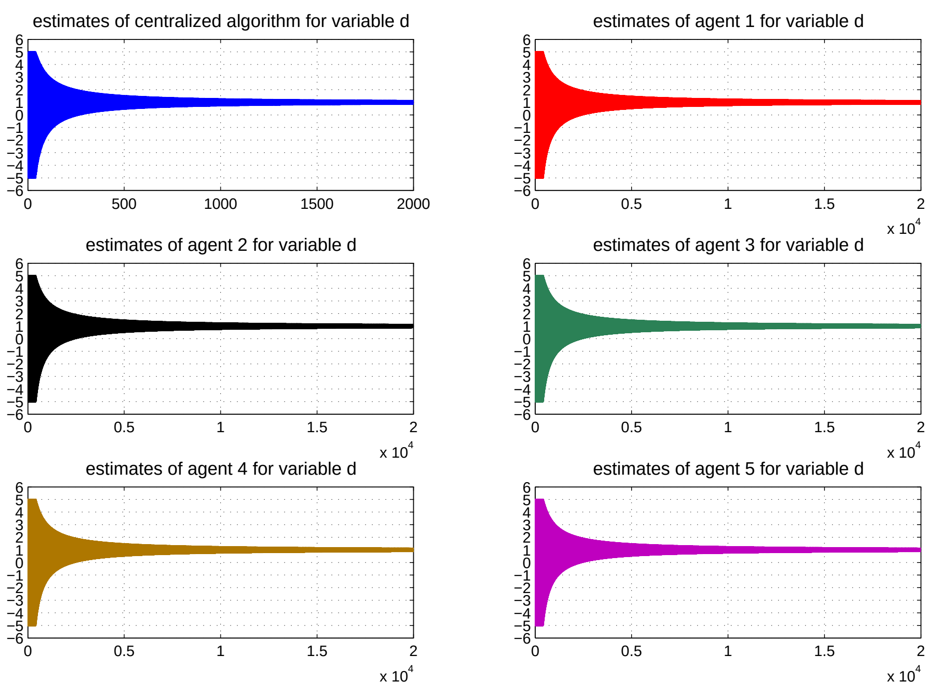


Fig. 9. Estimates of variable d in the DPPDS algorithm

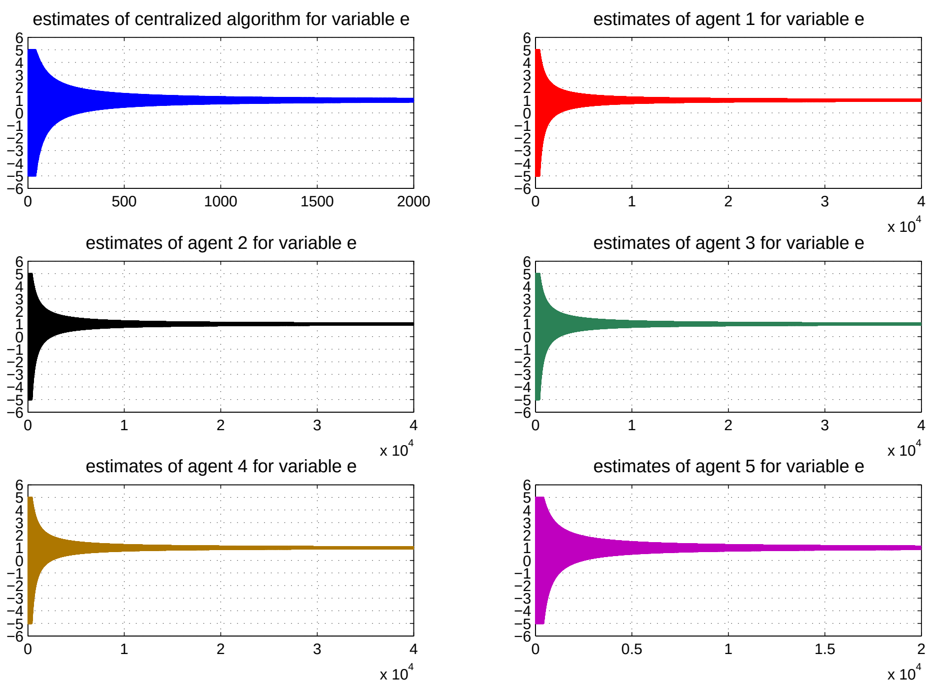
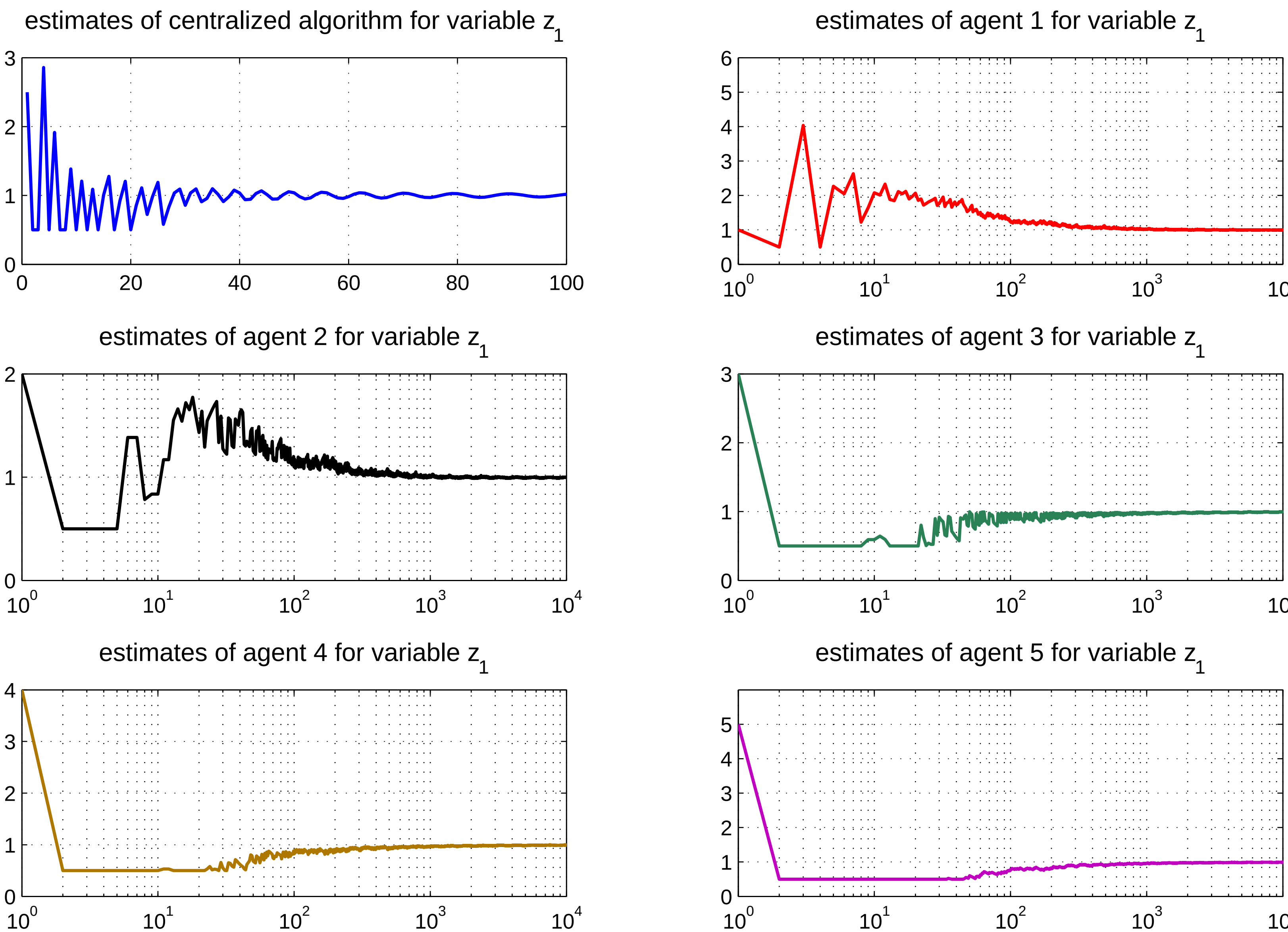
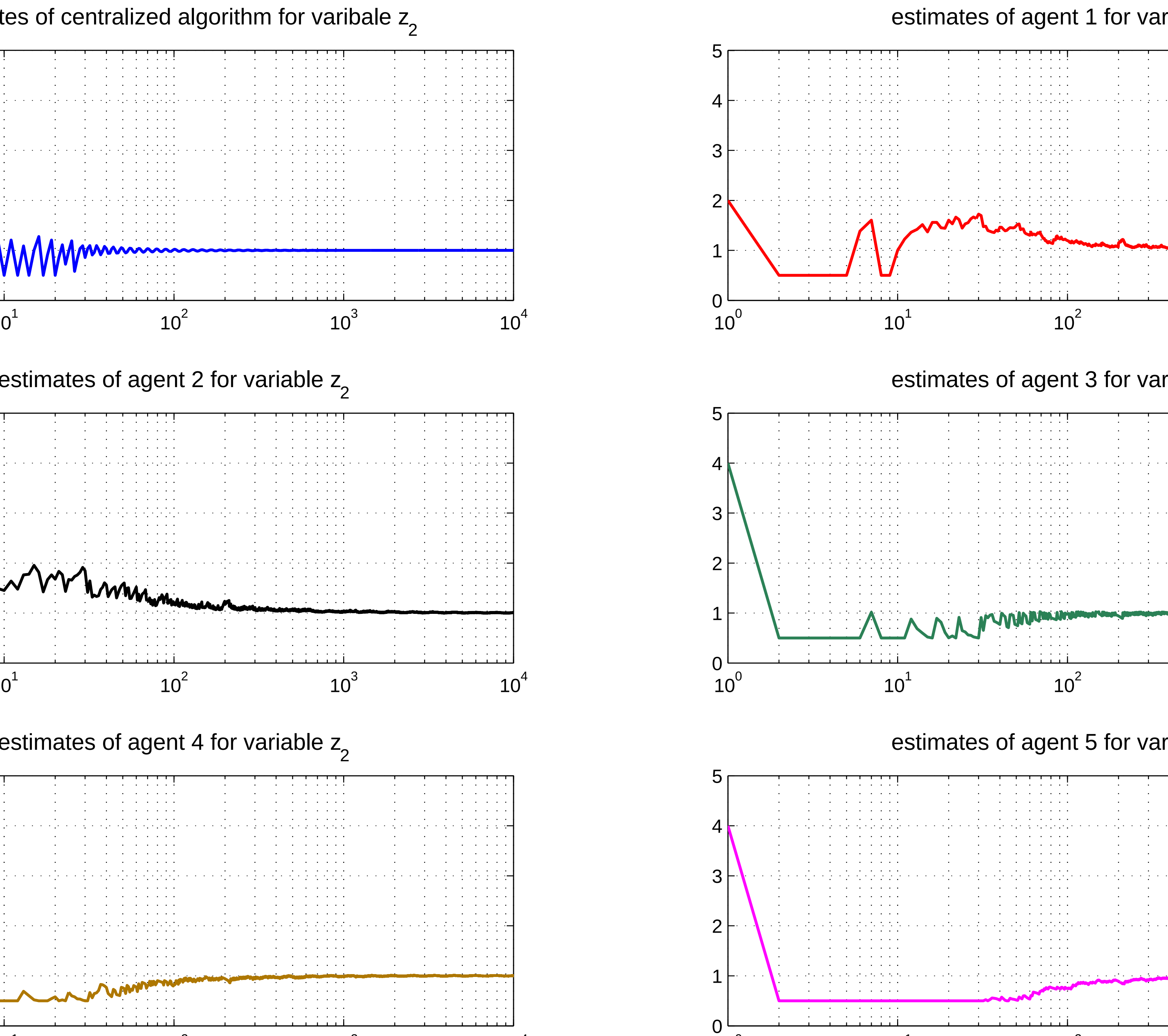


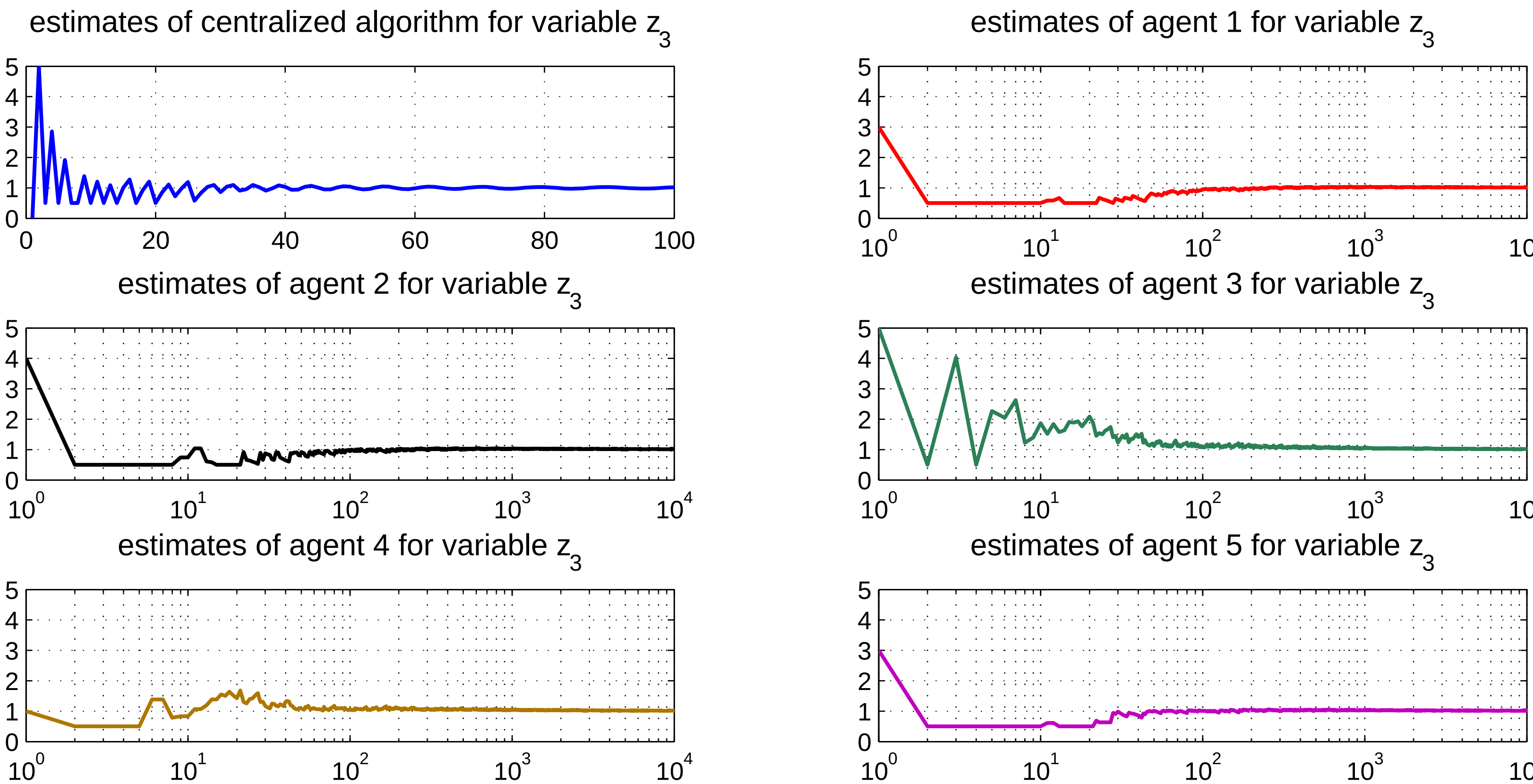
Fig. 10. Estimates of variable e in the DPPDS algorithm



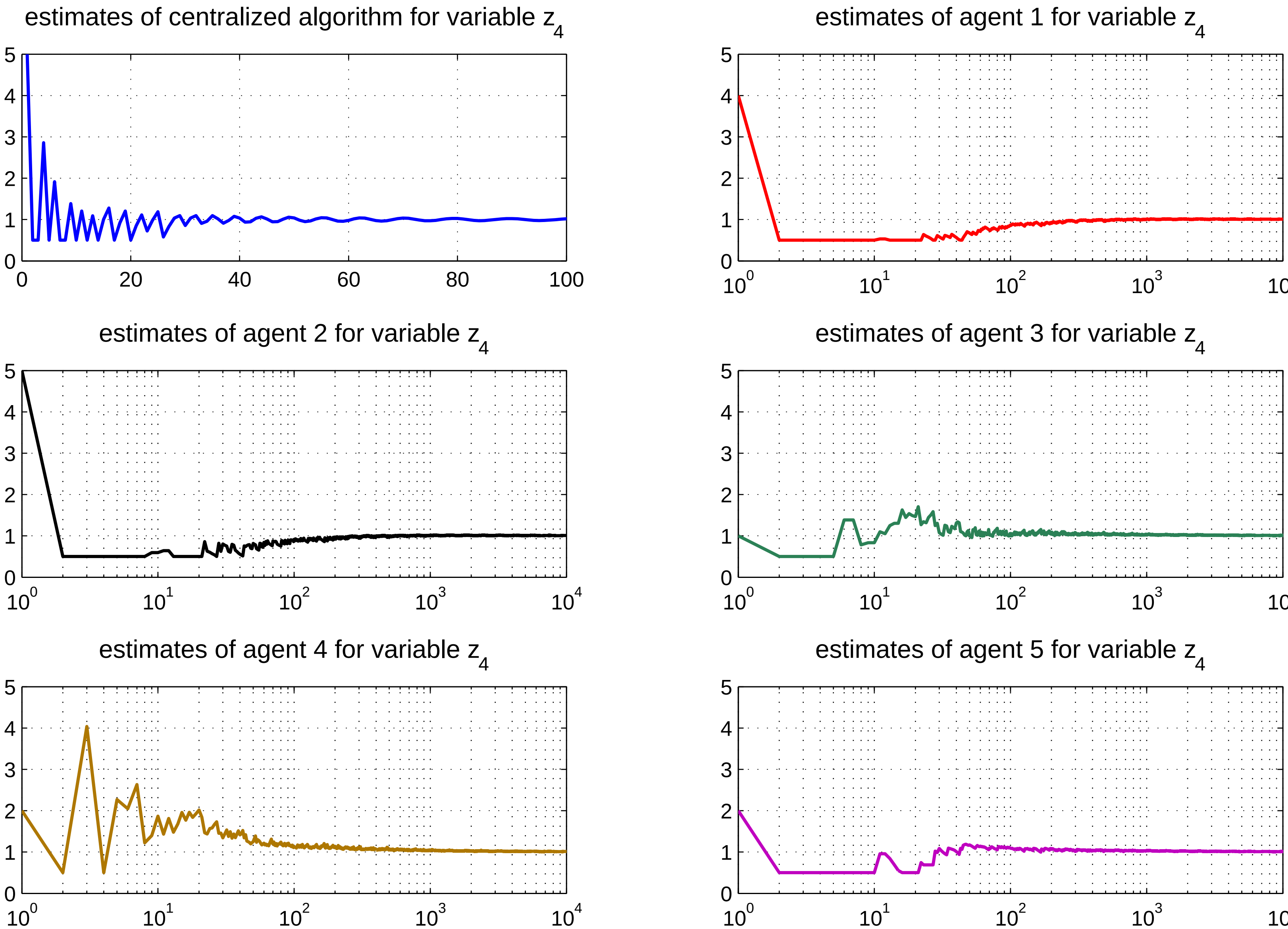
estimates of centralized algorithm for variable z_1
estimates of agent 1 for variable z_1
estimates of agent 2 for variable z_1
estimates of agent 3 for variable z_1
estimates of agent 4 for variable z_1
estimates of agent 5 for variable z_1
0 20 40 60 80 100
10^0 10^1 10^2 10^3 10^4

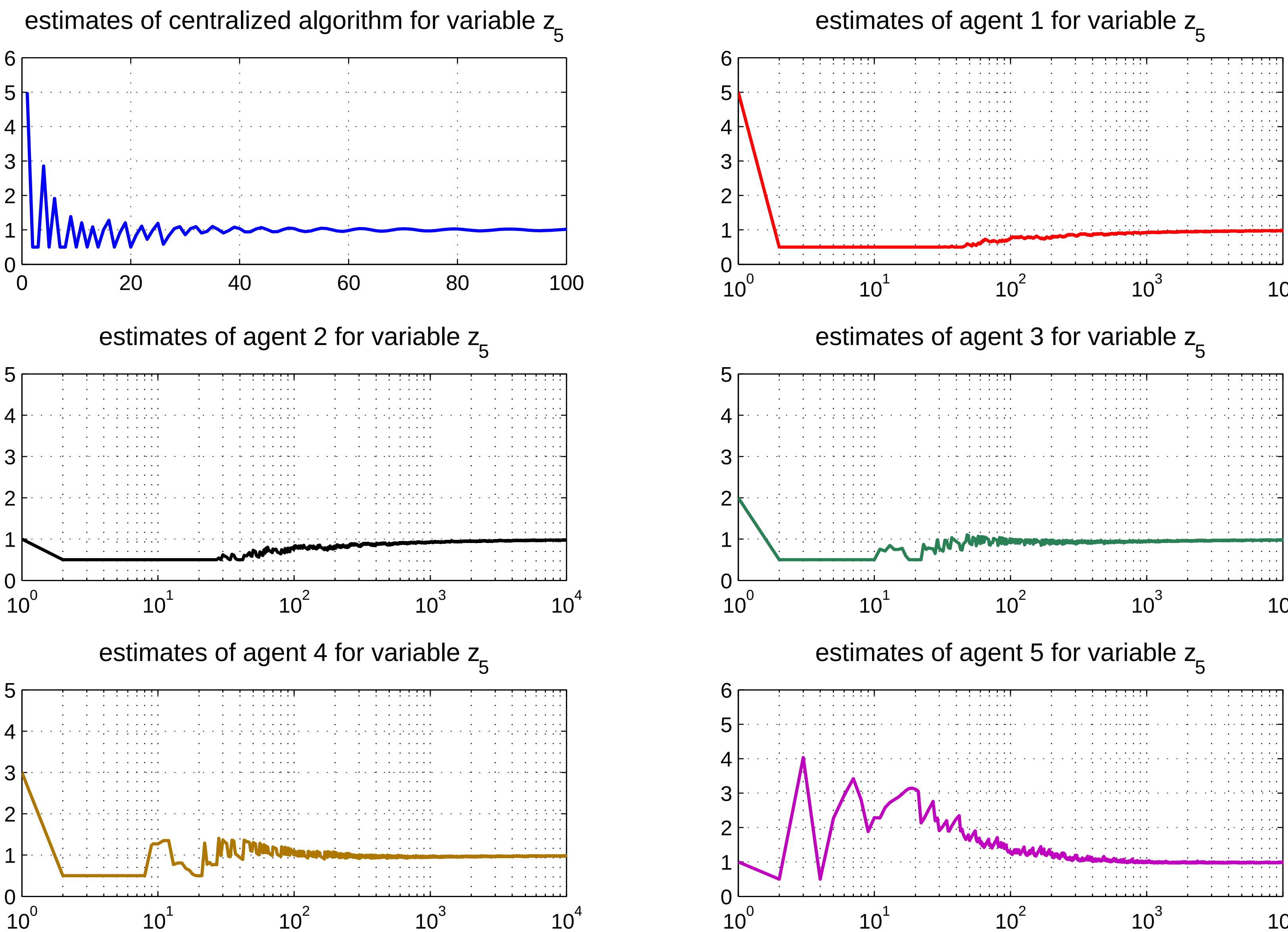


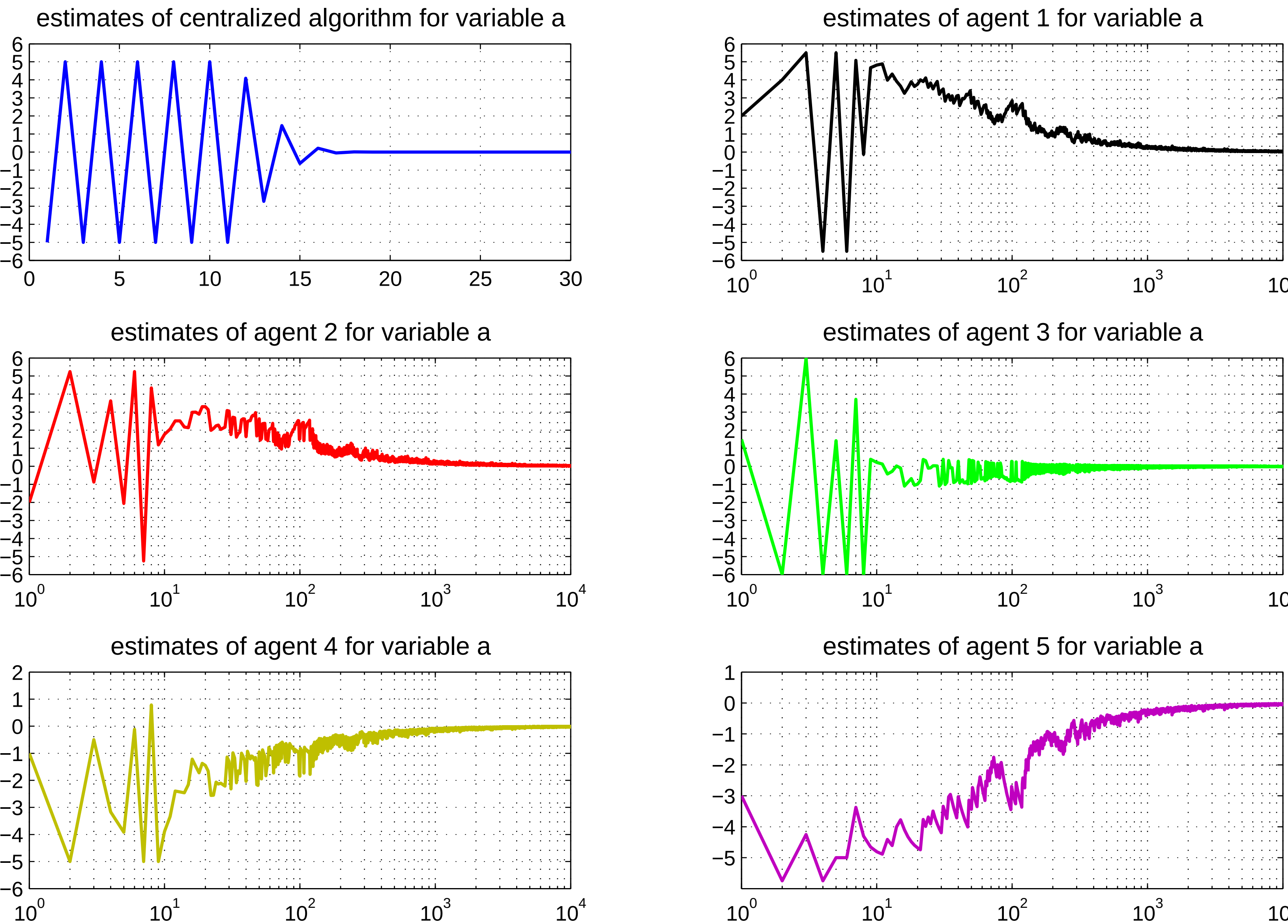
tes of centralized algorithm for varibale z_2
estimates of agent 1 for var
estimates of agent 2 for variable z_2
estimates of agent 3 for var
estimates of agent 4 for variable z_2
estimates of agent 5 for var
0
1
2
3
4
5
10^0
10^1
10^2
10^3
10^4

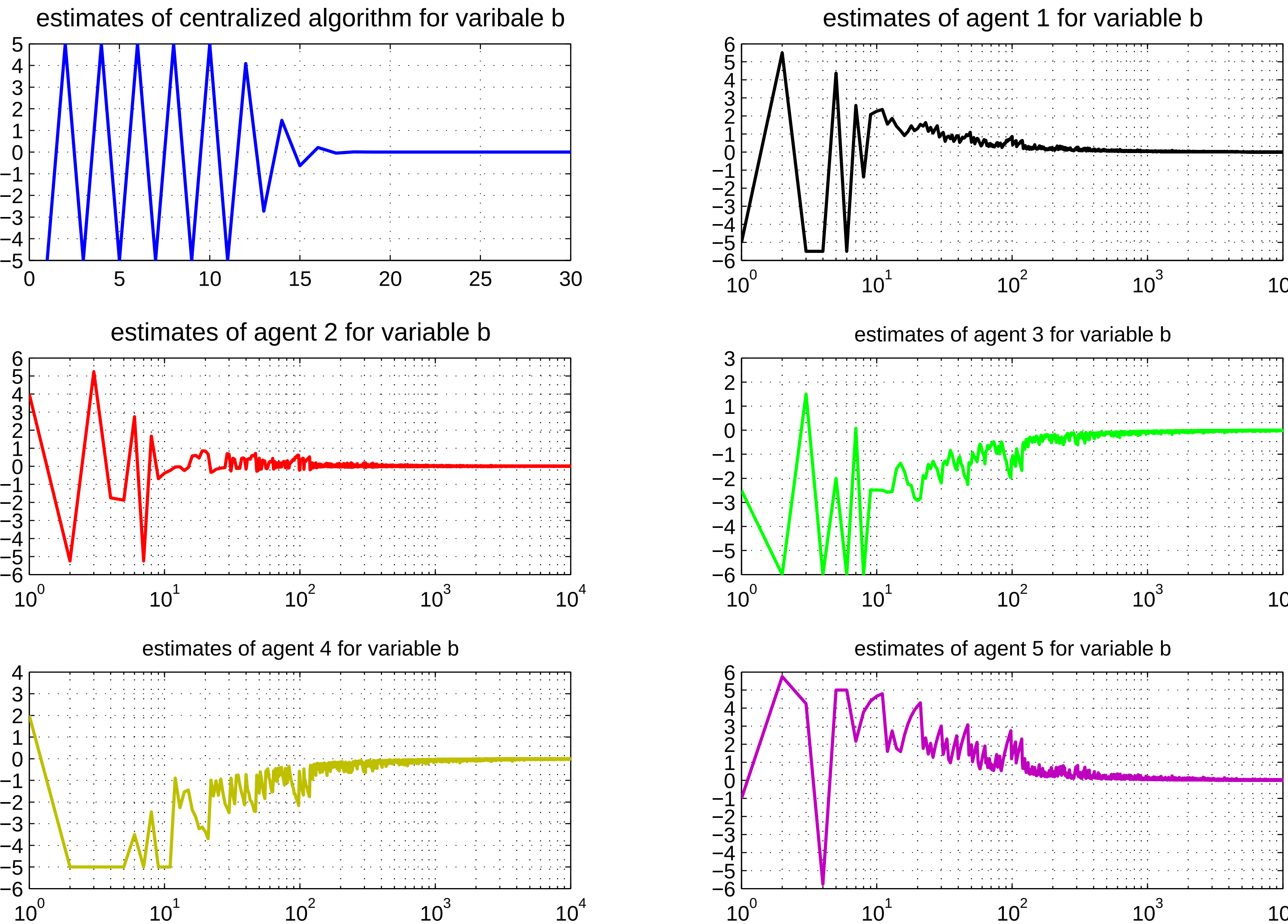


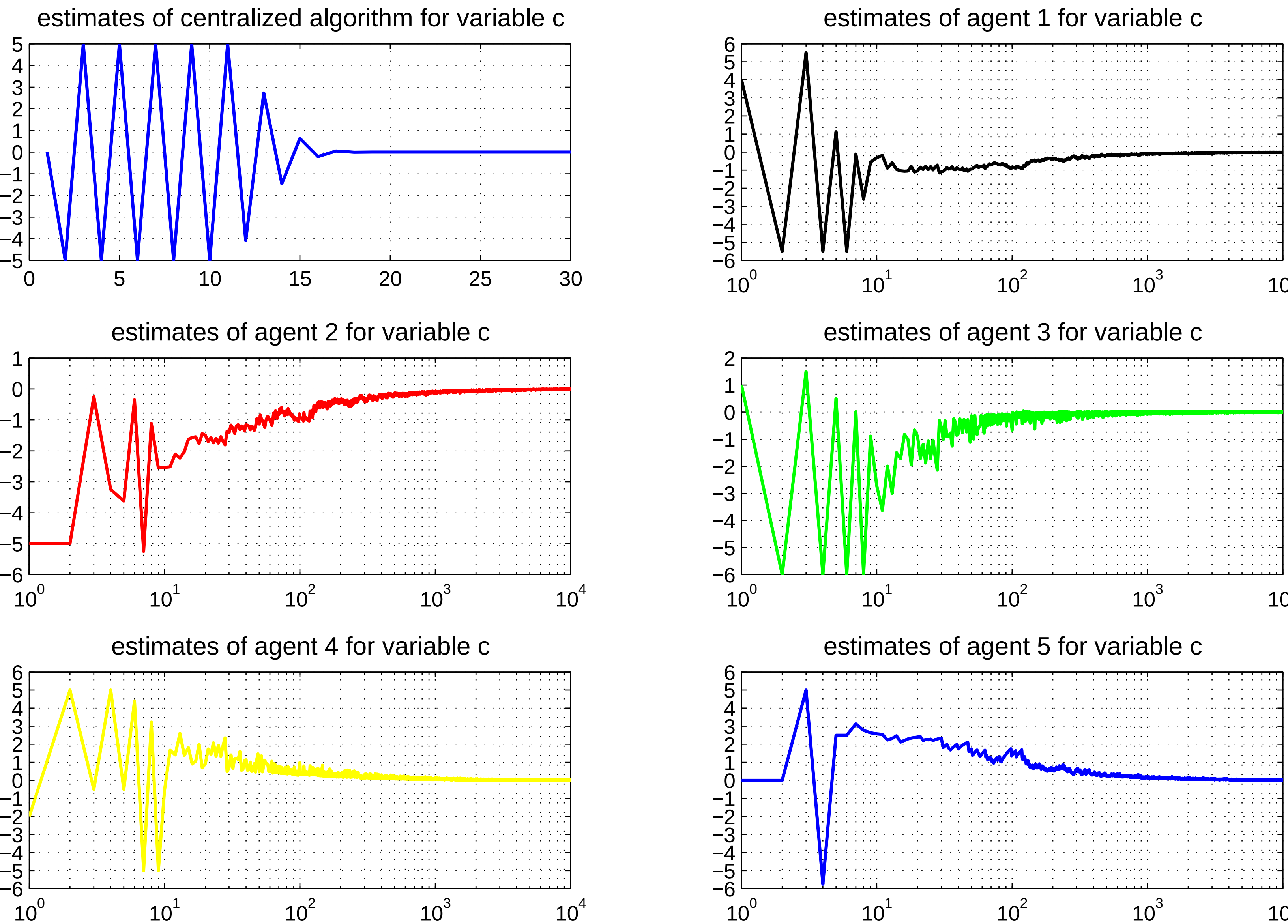
estimates of centralized algorithm for variable z3
estimates of agent 1 for variable z3
estimates of agent 2 for variable z3
estimates of agent 3 for variable z3
estimates of agent 4 for variable z3
estimates of agent 5 for variable z3


estimates of centralized algorithm for variable z_4
estimates of agent 1 for variable z_4
estimates of agent 2 for variable z_4
estimates of agent 3 for variable z_4
estimates of agent 4 for variable z_4
estimates of agent 5 for variable z_4
0
1
2
3
4
5
0
20
40
60
80
100
10^0
10^1
10^2
10^3
10^4

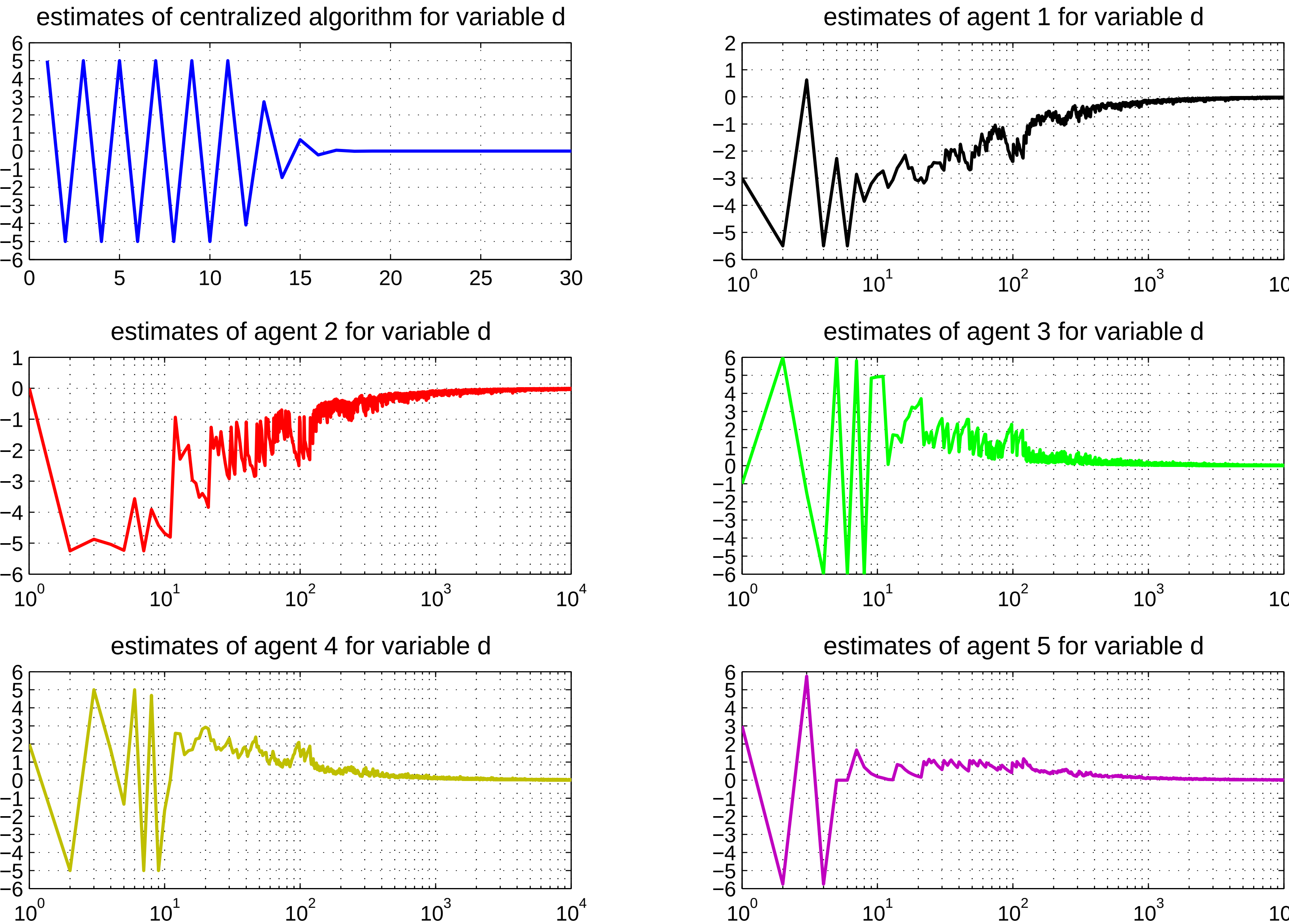

estimates of centralized algorithm for variable z5
0 1 2 3 4 5 6
0 20 40 60 80 100
estimates of agent 1 for variable z5
0 1 2 3 4 5 6
10^0 10^1 10^2 10^3 10
estimates of agent 2 for variable z5
0 1 2 3 4 5
10^0 10^1 10^2 10^3 10^4
estimates of agent 3 for variable z5
0 1 2 3 4 5
10^0 10^1 10^2 10^3 10
estimates of agent 4 for variable z5
0 1 2 3 4 5
10^0 10^1 10^2 10^3 10^4
estimates of agent 5 for variable z5
0 1 2 3 4 5 6
10^0 10^1 10^2 10^3 10

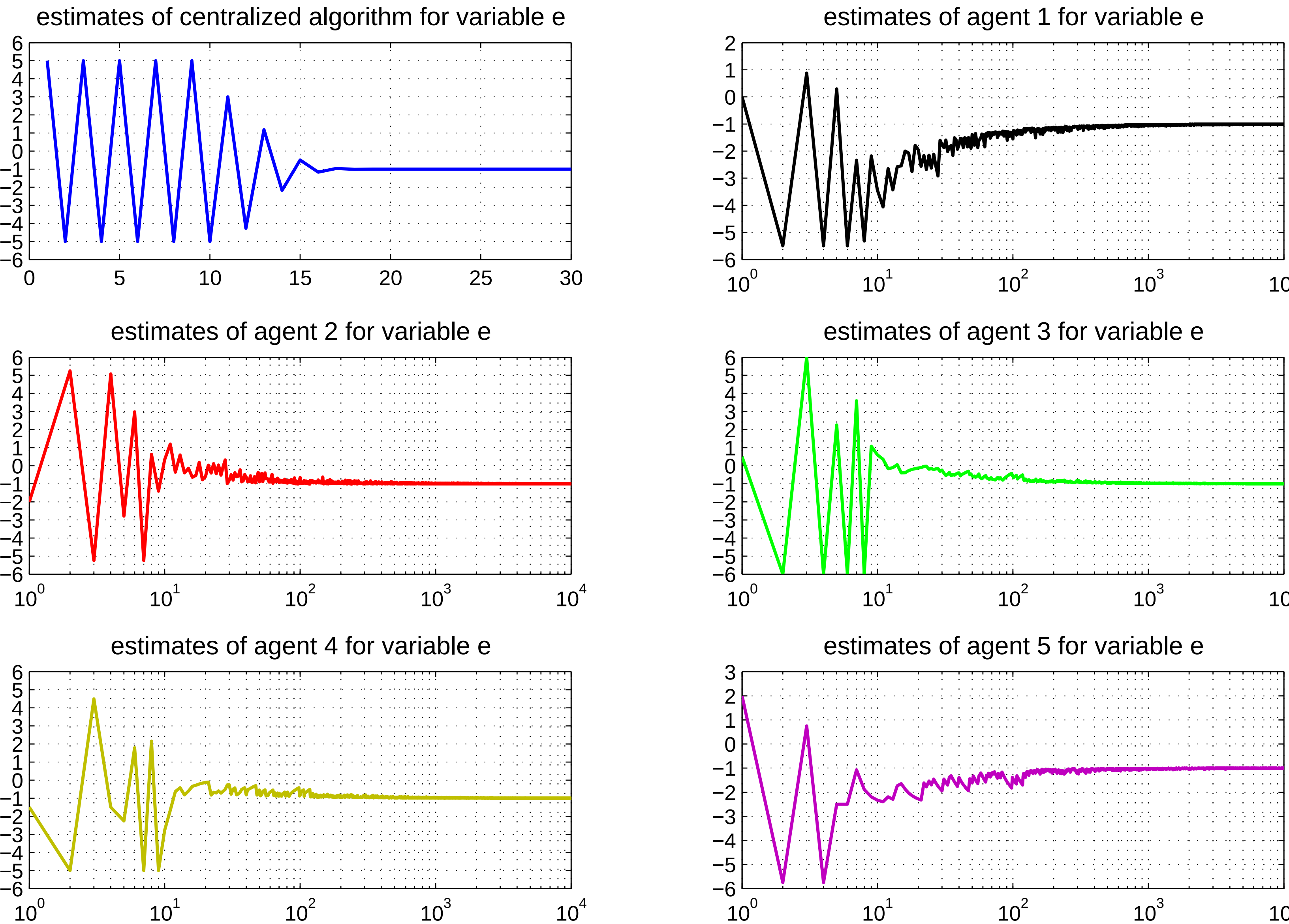

estimates of centralized algorithm for variable a
estimates of agent 1 for variable a
estimates of agent 2 for variable a
estimates of agent 3 for variable a
estimates of agent 4 for variable a
estimates of agent 5 for variable a
0 5 10 15 20 25 30
10^0 10^1 10^2 10^3 10^4


estimates of centralized algorithm for varibale b
estimates of agent 1 for variable b
estimates of agent 2 for variable b
estimates of agent 3 for variable b
estimates of agent 4 for variable b
estimates of agent 5 for variable b



estimates of centralized algorithm for variable c
estimates of agent 1 for variable c
estimates of agent 2 for variable c
estimates of agent 3 for variable c
estimates of agent 4 for variable c
estimates of agent 5 for variable c


estimates of centralized algorithm for variable d
estimates of agent 1 for variable d
estimates of agent 2 for variable d
estimates of agent 3 for variable d
estimates of agent 4 for variable d
estimates of agent 5 for variable d


estimates of centralized algorithm for variable e
6 5 4 3 2 1 0 −1 −2 −3 −4 −5 −6
0 5 10 15 20 25 30
estimates of agent 1 for variable e
2 1 0 −1 −2 −3 −4 −5 −6
10^0 10^1 10^2 10^3 10
estimates of agent 2 for variable e
6 5 4 3 2 1 0 −1 −2 −3 −4 −5 −6
10^0 10^1 10^2 10^3 10^4
estimates of agent 3 for variable e
6 5 4 3 2 1 0 −1 −2 −3 −4 −5 −6
10^0 10^1 10^2 10^3 10
estimates of agent 4 for variable e
6 5 4 3 2 1 0 −1 −2 −3 −4 −5 −6
10^0 10^1 10^2 10^3 10^4
estimates of agent 5 for variable e
3 2 1 0 −1 −2 −3 −4 −5 −6
10^0 10^1 10^2 10^3 10